

WO02083837

Publication Title:

METHODS FOR IDENTIFYING SMALL MOLECULES THAT BIND SPECIFIC RNA STRUCTURAL MOTIFS

Abstract:

Abstract of WO02083837

The present invention relates to a method for screening and identifying test compounds that bind to a preselected target ribonucleic acid ("RNA"). Direct, non-competitive binding assays are advantageously used to screen bead-based libraries of compounds for those that selectively bind to a preselected target RNA. Binding of target RNA molecules to a particular test compound is detected using any physical method that measures the altered physical property of the target RNA bound to a test compound. The structure of the test compound attached to the labeled RNA is also determined. The methods used will depend, in part, on the nature of the library screened. The methods of the present invention provide a simple, sensitive assay for high-throughput screening of libraries of compounds to identify pharmaceutical leads. Data supplied from the esp@cenet database - Worldwide

Courtesy of <http://v3.espacenet.com>

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
24 October 2002 (24.10.2002)

PCT

(10) International Publication Number
WO 02/083837 A1

- (51) International Patent Classification⁷: C12M 1/38, 1/40, C12Q 1/68
- (21) International Application Number: PCT/US02/11758
- (22) International Filing Date: 11 April 2002 (11.04.2002)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data: 60/282,966 11 April 2001 (11.04.2001) US
- (71) Applicant (for all designated States except US): PTC THERAPEUTICS, INC. [US/US]; 100 Corporate Court, Middlesex Business Center, South Plainfield, NJ 07080 (US).
- (72) Inventor; and
- (75) Inventor/Applicant (for US only): ALMSTEAD, Neil, G. [US/US]; 1 Crocus Drive, Holmdel, NJ 07733 (US).
- (74) Agents: CORUZZI, Laura, A. et al.; Pennie & Edmonds LLP, 1155 Avenue of the Americas, New York, NY 10036 (US).
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

- with international search report
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



A1

WO 02/083837

(54) Title: METHODS FOR IDENTIFYING SMALL MOLECULES THAT BIND SPECIFIC RNA STRUCTURAL MOTIFS

(57) Abstract: The present invention relates to a method for screening and identifying test compounds that bind to a preselected target ribonucleic acid ("RNA"). Direct, non-competitive binding assays are advantageously used to screen bead-based libraries of compounds for those that selectively bind to a preselected target RNA. Binding of target RNA molecules to a particular test compound is detected using any physical method that measures the altered physical property of the target RNA bound to a test compound. The structure of the test compound attached to the labeled RNA is also determined. The methods used will depend, in part, on the nature of the library screened. The methods of the present invention provide a simple, sensitive assay for high-throughput screening of libraries of compounds to identify pharmaceutical leads.

**METHODS FOR IDENTIFYING SMALL MOLECULES
THAT BIND SPECIFIC RNA STRUCTURAL MOTIFS**

5 This application claims the benefit of U.S. Provisional Application No. 60/282,966, filed April 11, 2001, which is incorporated herein by reference in its entirety.

1. INTRODUCTION

10 The present invention relates to a method for screening and identifying test compounds that bind to a preselected target ribonucleic acid ("RNA"). Direct, non-competitive binding assays are advantageously used to screen bead-based libraries of compounds for those that selectively bind to a preselected target RNA. Binding of target RNA molecules to a particular test compound is detected using any method that measures the altered physical property of the target RNA bound to a test compound. The methods of 15 the present invention provide a simple, sensitive assay for high-throughput screening of libraries of compounds to identify pharmaceutical leads.

2. BACKGROUND OF THE INVENTION

20 Protein-nucleic acid interactions are involved in many cellular functions, including transcription, RNA splicing, mRNA decay, and mRNA translation. Readily accessible synthetic molecules that can bind with high affinity to specific sequences of single- or double-stranded nucleic acids have the potential to interfere with these interactions in a controllable way, making them attractive tools for molecular biology and medicine. Successful approaches for blocking function of target nucleic acids include using duplex-forming antisense oligonucleotides (Miller, 1996, *Progress in Nucl. Acid Res. & Mol. Biol.* 52:261-291; Ojwang & Rando, 1999, *Achieving antisense inhibition by oligodeoxynucleotides containing N, modified 2'-deoxyguanosine using tumor necrosis factor receptor type 1, METHODS: A Companion to Methods in Enzymology* 18:244-251) and peptide nucleic acids ("PNA") (Nielsen, 1999, *Current Opinion in Biotechnology* 10:71-75), which bind to nucleic acids via Watson-Crick base-pairing. Triplex-forming 25 anti-gene oligonucleotides can also be designed (Ping *et al.*, 1997, *RNA* 3:850-860; Aggarwal *et al.*, 1996, *Cancer Res.* 56:5156-5164; U.S. Patent No. 5,650,316), as well as pyrrole-imidazole polyamide oligomers (Gottesfeld *et al.*, 1997, *Nature* 387:202-205; White *et al.*, 1998, *Nature* 391:468-471), which are specific for the major and minor grooves of a 30 double helix, respectively.

In addition to synthetic nucleic acids (*i.e.*, antisense, ribozymes, and triplex-forming molecules), there are examples of natural products that interfere with deoxyribonucleic acid ("DNA") or RNA processes such as transcription or translation. For example, certain carbohydrate-based host cell factors, calicheamicin oligosaccharides, interfere with the sequence-specific binding of transcription factors to DNA and inhibit transcription *in vivo* (Ho *et al.*, 1994, Proc. Natl. Acad. Sci. USA 91:9203-9207; Liu *et al.*, 1996, Proc. Natl. Acad. Sci. USA 93:940-944). Certain classes of known antibiotics have been characterized and were found to interact with RNA. For example, the antibiotic thiostreptone binds tightly to a 60-mer from ribosomal RNA (Cundliffe *et al.*, 1990, in *The Ribosome: Structure, Function & Evolution* (Schlessinger *et al.*, eds.) American Society for Microbiology, Washington, D.C. pp. 479-490). Bacterial resistance to various antibiotics often involves methylation at specific rRNA sites (Cundliffe, 1989, Ann. Rev. Microbiol. 43:207-233). Aminoglycosidic aminocyclitol (aminoglycoside) antibiotics and peptide antibiotics are known to inhibit group I intron splicing by binding to specific regions of the RNA (von Ahsen *et al.*, 1991, Nature (London) 353:368-370). Some of these same aminoglycosides have also been found to inhibit hammerhead ribozyme function (Stage *et al.*, 1995, RNA 1:95-101). In addition, certain aminoglycosides and other protein synthesis inhibitors have been found to interact with specific bases in 16S rRNA (Woodcock *et al.*, 1991, EMBO J. 10:3099-3103). An oligonucleotide analog of the 16S rRNA has also been shown to interact with certain aminoglycosides (Purohit *et al.*, 1994, Nature 370:659-662). A molecular basis for hypersensitivity to aminoglycosides has been found to be located in a single base change in mitochondrial rRNA (Hutchin *et al.*, 1993, Nucleic Acids Res. 21:4174-4179). Aminoglycosides have also been shown to inhibit the interaction between specific structural RNA motifs and the corresponding RNA binding protein. Zapp *et al.* (Cell, 1993, 74:969-978) has demonstrated that the aminoglycosides neomycin B, lividomycin A, and tobramycin can block the binding of Rev, a viral regulatory protein required for viral gene expression, to its viral recognition element in the IIB (or RRE) region of HIV RNA. This blockage appears to be the result of competitive binding of the antibiotics directly to the RRE RNA structural motif.

Single stranded sections of RNA can fold into complex tertiary structures consisting of local motifs such as loops, bulges, pseudoknots, guanosine quartets and turns (Chastain & Tinoco, 1991, Progress in Nucleic Acid Res. & Mol. Biol. 41:131-177; Chow & Bogdan, 1997, Chemical Reviews 97:1489-1514; Rando & Hogan, 1998, Biologic activity of guanosine quartet forming oligonucleotides in "Applied Antisense Oligonucleotide Technology" Stein. & Krieg (eds) John Wiley and Sons, New York, pages 335-352). Such

structures can be critical to the activity of the nucleic acid and affect functions such as regulation of mRNA transcription, stability, or translation (Weeks & Crothers, 1993, *Science* 261:1574-1577). The dependence of these functions on the native three-dimensional structural motifs of single-stranded stretches of nucleic acids makes it difficult to identify or design synthetic agents that bind to these motifs using general, simple-to-use sequence-specific recognition rules for the formation of double- and triple-helical nucleic acids used in the design of antisense and ribozyme type molecules. Approaches to screening generally involve competitive assays designed to identify compounds that disrupt the interaction between a target RNA and a physiological, host cell factor(s) that had been previously identified to specifically interact with that particular target RNA. In general, such assays require the identification and characterization of the host cell factor(s) deemed to be required for the function of the target RNA. Both the target RNA and its preselected host cell binding partner are used in a competitive format to identify compounds that disrupt or interfere with the two components in the assay.

Citation or identification of any reference in Section 2 of this application is not an admission that such reference is available as prior art to the present invention.

3. SUMMARY OF THE INVENTION

The present invention relates to methods for identifying compounds that bind to preselected target elements of nucleic acids including, but not limited to, specific RNA sequences, RNA structural motifs, and/or RNA structural elements. The specific target RNA sequences, RNA structural motifs, and/or RNA structural elements are used as targets for screening small molecules and identifying those that directly bind these specific sequences, motifs, and/or structural elements. For example, methods are described in which a preselected target RNA having a detectable label is used to screen a library of test compounds, preferably under physiologic conditions. Any complexes formed between the target RNA and a member of the library are identified using methods that detect the labeled target RNA bound to a test compound. In particular, the present invention relates to methods for using a target RNA having a detectable label to screen a bead-based library of test compounds. Compounds in the bead-based library that bind to the labeled target RNA will form a bead-based detectably labeled complex, which can be separated from the unbound beads and unbound target RNA in the liquid phase by a number of physical means, including, but not limited to, flow cytometry, affinity chromatography, manual batch mode separation, suspension of beads in electric fields, and microwave of the bead-based detectably labeled complex. The detectably labeled complex can then be identified by the label on the target

RNA and removed from the uncomplexed, unlabeled test compounds in the library. The structure of the test compound complexed with the labeled RNA is then ascertained by *de novo* structure determination of the test compounds using, for example, mass spectrometry or nuclear magnetic resonance ("NMR"). The test compounds identified are useful for any purpose to which a binding reaction may be put, for example in assay methods, diagnostic procedures, cell sorting, as inhibitors of target molecule function, as probes, as sequestering agents and the like. In addition, small organic molecules which interact specifically with target RNA molecules may be useful as lead compounds for the development of therapeutic agents.

The methods described herein for the identification of compounds that directly bind to a particular preselected target RNA are well suited for high-throughput screening. The direct binding method of the invention offers advantages over drug screening systems for competitors that inhibit the formation of naturally-occurring RNA binding protein:target RNA complexes; *i.e.*, competitive assays. The direct binding method of the invention is rapid and can be set up to be readily performed, *e.g.*, by a technician, making it amenable to high throughput screening. The method of the invention also eliminates the bias inherent in the competitive drug screening systems, which require the use of a preselected host cell factor that may not have physiological relevance to the activity of the target RNA. Instead, the methods of the invention are used to identify any compound that can directly bind to specific target RNA sequences, RNA structural motifs, and/or RNA structural elements, preferably under physiologic conditions. As a result, the compounds so identified can inhibit the interaction of the target RNA with any one or more of the native host cell factors (whether known or unknown) required for activity of the RNA *in vivo*.

The present invention may be understood more fully by reference to the detailed description and examples, which are intended to illustrate non-limiting embodiments of the invention.

3.1. Definitions

As used herein, a "target nucleic acid" refers to RNA, DNA, or a chemically modified variant thereof. In a preferred embodiment, the target nucleic acid is RNA. A target nucleic acid also refers to tertiary structures of the nucleic acids, such as, but not limited to loops, bulges, pseudoknots, guanosine quartets and turns. A target nucleic acid also refers to RNA elements such as, but not limited to, the HIV TAR element, internal ribosome entry site, "slippery site", instability elements, and adenylate uridylate-rich

elements, which are described in Section 4.1. Non-limiting examples of target nucleic acids are presented in Section 4.1 and Section 5.

As used herein, a "library" refers to a plurality of test compounds with which a target nucleic acid molecule is contacted. A library can be a combinatorial library, *e.g.*, a collection of test compounds synthesized using combinatorial chemistry techniques, or a collection of unique chemicals of low molecular weight (less than 1000 daltons) that each occupy a unique three-dimensional space.

As used herein, a "label" or "detectable label" is a composition that is detectable, either directly or indirectly, by spectroscopic, photochemical, biochemical, immunochemical, or chemical means. For example, useful labels include radioactive isotopes (*e.g.*, ^{32}P , ^{35}S , and ^3H), dyes, fluorescent dyes, electron-dense reagents, enzymes and their substrates (*e.g.*, as commonly used in enzyme-linked immunoassays, *e.g.*, alkaline phosphatase and horse radish peroxidase), biotin, digoxigenin, or haptens and proteins for which antisera or monoclonal antibodies are available. Moreover, a label or detectable moiety can include an "affinity tag" that, when coupled with the target nucleic acid and incubated with a test compound or compound library, allows for the affinity capture of the target nucleic acid along with molecules bound to the target nucleic acid. One skilled in the art will appreciate that a affinity tag bound to the target nucleic acids has, by definition, a complimentary ligand coupled to a solid support that allows for its capture. For example, useful affinity tags and complimentary ligands include, but are not limited to, biotin-streptavidin, complimentary nucleic acid fragments (*e.g.*, oligo dT-oligo dA, oligo T-oligo A, oligo dG-oligo dC, oligo G-oligo C), aptamer complexes, or haptens and proteins for which antisera or monoclonal antibodies are available. The label or detectable moiety is typically bound, either covalently, through a linker or chemical bound, or through ionic, van der Waals or hydrogen bonds to the molecule to be detected.

As used herein, a "dye" refers to a molecule that, when exposed to radiation, emits radiation at a level that is detectable visually or via conventional spectroscopic means. As used herein, a "visible dye" refers to a molecule having a chromophore that absorbs radiation in the visible region of the spectrum (*i.e.*, having a wavelength of between about 400 nm and about 700 nm) such that the transmitted radiation is in the visible region and can be detected either visually or by conventional spectroscopic means. As used herein, an "ultraviolet dye" refers to a molecule having a chromophore that absorbs radiation in the ultraviolet region of the spectrum (*i.e.*, having a wavelength of between about 30 nm and about 400 nm). As used herein, an "infrared dye" refers to a molecule having a chromophore that absorbs radiation in the infrared region of the spectrum (*i.e.*, having a wavelength

between about 700 nm and about 3,000 nm). A "chromophore" is the network of atoms of the dye that, when exposed to radiation, emits radiation at a level that is detectable visually or via conventional spectroscopic means. One of skill in the art will readily appreciate that although a dye absorbs radiation in one region of the spectrum, it may emit radiation in another region of the spectrum. For example, an ultraviolet dye may emit radiation in the visible region of the spectrum. One of skill in the art will also readily appreciate that a dye can transmit radiation or can emit radiation via fluorescence or phosphorescence.

The phrase "pharmaceutically acceptable salt(s)," as used herein includes but is not limited to salts of acidic or basic groups that may be present in test compounds identified using the methods of the present invention. Test compounds that are basic in nature are capable of forming a wide variety of salts with various inorganic and organic acids. The acids that can be used to prepare pharmaceutically acceptable acid addition salts of such basic compounds are those that form non-toxic acid addition salts, *i.e.*, salts containing pharmacologically acceptable anions, including but not limited to sulfuric, citric, maleic, acetic, oxalic, hydrochloride, hydrobromide, hydroiodide, nitrate, sulfate, bisulfate, phosphate, acid phosphate, isonicotinate, acetate, lactate, salicylate, citrate, acid citrate, tartrate, oleate, tannate, pantothenate, bitartrate, ascorbate, succinate, maleate, gentisinate, fumarate, gluconate, glucaronate, saccharate, formate, benzoate, glutamate, methanesulfonate, ethanesulfonate, benzenesulfonate, p-toluenesulfonate and pamoate (*i.e.*, 1,1'-methylene-bis-(2-hydroxy-3-naphthoate)) salts. Test compounds that include an amino moiety may form pharmaceutically or cosmetically acceptable salts with various amino acids, in addition to the acids mentioned above. Test compounds that are acidic in nature are capable of forming base salts with various pharmacologically or cosmetically acceptable cations. Examples of such salts include alkali metal or alkaline earth metal salts and, particularly, calcium, magnesium, sodium lithium, zinc, potassium, and iron salts.

By "substantially one type of test compound," as used herein, is meant that the assay can be performed in such a fashion that at some point, only one compound need be used in each reaction so that, if the result is indicative of a binding event occurring between the target RNA molecule and the test compound the test compound, can be easily identified.

4. DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to methods for identifying compounds that bind to preselected target elements of nucleic acids, in particular, RNAs, including but not limited to preselected target RNA sequencing structural motifs, or structural elements. Methods are described in which a preselected target RNA having a detectable label is used to screen a

library of test compounds. Any complexes formed between the target RNA and a member of the library are identified using methods that detect the labeled target RNA bound to a test compound. In particular, the present invention relates to methods for using a target RNA having a detectable label to screen a bead-based library of test compounds. Compounds in the bead-based library that bind to the labeled target RNA will form a bead-based detectably labeled complex, which can be separated from the unbound target RNA in the liquid phase by a number of physical means, such as, but not limited to, flow cytometry, affinity chromatography, manual batch mode separation, suspension of beads in electric fields, and microwave of the bead-based detectably labeled complex. The detectably labeled complex can then be identified by the label on the target RNA and removed from the uncomplexed, unlabeled test compounds in the library. The structure of the test compound attached to the labeled RNA is then ascertained by *de novo* structure determination of the test compounds using, for example, mass spectrometry or nuclear magnetic resonance ("NMR").

Thus, the methods of the present invention provide a simple, sensitive assay for high-throughput screening of libraries of test compounds, in which the test compounds of the library that specifically bind a preselected target nucleic acid are easily distinguished from non-binding members of the library. The structures of the binding molecules are ascertained by *de novo* structure determination of the test compounds using, for example, mass spectrometry or nuclear magnetic resonance ("NMR"). The test compounds so identified are useful for any purpose to which a binding reaction may be put, for example in assay methods, diagnostic procedures, cell sorting, as inhibitors of target molecule function, as probes, as sequestering agents and lead compounds for development of therapeutics, and the like. Small organic compounds that are identified to interact specifically with the target RNA molecules are particularly attractive candidates as lead compounds for the development of therapeutic agents.

The assay of the invention reduces bias introduced by competitive binding assays which require the identification and use of a host cell factor (presumably essential for modulating RNA function) as a binding partner for the target RNA. The assays of the present invention are designed to detect any compound or agent that binds to the target RNA, preferably under physiologic conditions. Such agents can then be tested for biological activity, without establishing or guessing which host cell factor or factors is required for modulating the function and/or activity of the target RNA.

Section 4.1 describes examples of protein-RNA interactions that are important in a variety of cellular functions and several target RNA elements that can be used to identify test compounds. Compounds that inhibit these interactions by binding to the RNA and

successfully competing with the natural protein or host cell factor that endogenously binds to the RNA may be important, *e.g.*, in treating or preventing a disease or abnormal condition, such as an infection or unchecked growth. Section 4.2 describes detectable labels for target nucleic acids that are useful in the methods of the invention. Section 4.3 describes libraries of test compounds. Section 4.4 provides conditions for binding a labeled target RNA to a test compound of a library and detecting RNA binding to a test compound using the methods of the invention. Section 4.5 provides methods for separating complexes of target RNAs bound to a test compound from an unbound RNA. Section 4.6 describes methods for identifying test compounds that are bound to the target RNA. Section 4.7 describes a secondary, biological screen of test compounds identified by the methods of the invention to test the effect of the test compounds *in vivo*. Section 4.8 describes the use of test compounds identified by the methods of the invention for treating or preventing a disease or abnormal condition in mammals.

15

4.1. Biologically Important RNA-Host Cell Factor Interactions

Nucleic acids, and in particular RNAs, are capable of folding into complex tertiary structures that include bulges, loops, triple helices and pseudoknots, which can provide binding sites for host cell factors, such as proteins and other RNAs. RNA-protein and RNA-RNA interactions are important in a variety cellular functions, including transcription, RNA splicing, RNA stability and translation. Furthermore, the binding of such host cell factors to RNAs may alter the stability and translational efficiency of such RNAs, and accordingly affect subsequent translation. For example, some diseases are associated with protein overproduction or decreased protein function. In this case, the identification of compounds to modulate RNA stability and translational efficiency will be useful to treat and prevent such diseases.

The methods of the present invention are useful for identifying test compounds that bind to target RNA elements in a high throughput screening assay of libraries of test compounds in solution. In particular, the methods of the present invention are useful for identifying a test compound that binds to a target RNA elements and inhibits the interaction of that RNA with one or more host cell factors *in vivo*. The molecules identified using the methods of the invention are useful for inhibiting the formation of a specific bound RNA:host cell factor complexes *in vivo*.

In some embodiments, test compounds identified by the methods of the invention are useful for increasing or decreasing the translation of messenger RNAs ("mRNAs"), *e.g.*, protein production, by binding to one or more regulatory elements in the 5'

untranslated region, the 3' untranslated region, or the coding region of the mRNA. Compounds that bind to mRNA can, *inter alia*, increase or decrease the rate of mRNA processing, alter its transport through the cell, prevent or enhance binding of the mRNA to 5 ribosomes, suppressor proteins or enhancer proteins, or alter mRNA stability. Accordingly, compounds that increase or decrease mRNA translation can be used to treat or prevent disease. For example, diseases associated with protein overproduction, such as amyloidosis, or with the production of mutant proteins, such as *Ras*, can be treated or prevented by 10 decreasing translation of the mRNA that codes for the overproduced protein, thus inhibiting production of the protein. Conversely, the symptoms of diseases associated with decreased protein function, such as hemophilia, may be treated by increasing translation of mRNA coding for the protein whose function is decreased, *e.g.*, factor IX in some forms of hemophilia.

The methods of the invention can be used to identify compounds that bind to 15 mRNAs coding for a variety of proteins with which the progression of diseases in mammals is associated. These mRNAs include, but are not limited to, those coding for amyloid protein and amyloid precursor protein; anti-angiogenic proteins such as angiotatin, endostatin, METH-1 and METH-2; apoptosis inhibitor proteins such as survivin, clotting factors such as Factor IX, Factor VIII, and others in the clotting cascade; collagens; cyclins and cyclin 20 inhibitors, such as cyclin dependent kinases, cyclin D1, cyclin E, WAF1, cdk4 inhibitor, and MTS1; cystic fibrosis transmembrane conductance regulator gene (CFTR); cytokines such as IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, IL-14, IL-15, IL-16, IL-17 and other interleukins; hematopoietic growth factors such as erythropoietin (Epo); colony stimulating factors such as G-CSF, GM-CSF, M-CSF, SCF and 25 thrombopoietin; growth factors such as BDNF, BMP, GGRP, EGF, FGF, GDNF, GGF, HGF, IGF-1, IGF-2, KGF, myotrophin, NGF, OSM, PDGF, somatotrophin, TGF- β , TGF- α and VEGF; antiviral cytokines such as interferons, antiviral proteins induced by interferons, TNF- α , and TNF- β ; enzymes such as cathepsin K, cytochrome P-450 and other cytochromes, farnesyl transferase, glutathione-s transferases, heparanase, HMG CoA synthetase, N- 30 acetyltransferase, phenylalanine hydroxylase, phosphodiesterase, ras carboxyl-terminal protease, telomerase and TNF converting enzyme; glycoproteins such as cadherins, *e.g.*, N-cadherin and E-cadherin; cell adhesion molecules; selectins; transmembrane glycoproteins such as CD40; heat shock proteins; hormones such as 5- α reductase, atrial natriuretic factor, calcitonin, corticotrophin releasing factor, diuretic hormones, glucagon, gonadotropin, 35 gonadotropin releasing hormone, growth hormone, growth hormone releasing factor, somatotropin, insulin, leptin, luteinizing hormone, luteinizing hormone releasing hormone,

parathyroid hormone, thyroid hormone, and thyroid stimulating hormone; proteins involved in immune responses, including antibodies, CTLA4, hemagglutinin, MHC proteins, VLA-4, and kallikrein-kininogen-kinin system; ligands such as CD4; oncogene products such as *sis*, *hst*, protein tyrosine kinase receptors, *ras*, *abl*, *mos*, *myc*, *fos*, *jun*, *H-ras*, *ki-ras*, *c-fms*, *bcl-2*, *L-myc*, *c-myc*, *gip*, *gsp*, and *HER-2*; receptors such as bombesin receptor, estrogen receptor, GABA receptors, growth factor receptors including EGFR, PDGFR, FGFR, and NGFR, GTP-binding regulatory proteins, interleukin receptors, ion channel receptors, leukotriene receptor antagonists, lipoprotein receptors, opioid pain receptors, substance P receptors, 10 retinoic acid and retinoid receptors, steroid receptors, T-cell receptors, thyroid hormone receptors, TNF receptors; tissue plasminogen activator; transmembrane receptors; transmembrane transporting systems, such as calcium pump, proton pump, Na/Ca exchanger, MRP1, MRP2, P170, LRP, and cMOAT; transferrin; and tumor suppressor gene products such as *APC*, *brca1*, *brca2*, *DCC*, *MCC*, *MTS1*, *NF1*, *NF2*, *nm23*, *p53* and *Rb*. In addition to 15 the eukaryotic genes listed above, the invention, as described, can be used to define molecules that interrupt viral, bacterial or fungal transcription or translation efficiencies and therefore form the basis for a novel anti-infectious disease therapeutic. Other target genes include, but are not limited to, those disclosed in Section 4.1 and Section 5.

The methods of the invention can be used to identify mRNA-binding test 20 compounds for increasing or decreasing the production of a protein, thus treating or preventing a disease associated with decreasing or increasing the production of said protein, respectively. The methods of the invention may be useful for identifying test compounds for treating or preventing a disease in mammals, including cats, dogs, swine, horses, goats, sheep, cattle, primates and humans. Such diseases include, but are not limited to, 25 amyloidosis, hemophilia, Alzheimer's disease, atherosclerosis, cancer, giantism, dwarfism, hypothyroidism, hyperthyroidism, inflammation, cystic fibrosis, autoimmune disorders, diabetes, aging, obesity, neurodegenerative disorders, and Parkinson's disease. Other diseases include, but are not limited to, those described in Section 4.1 and diseases caused by aberrant expression of the genes disclosed in Example 5. In addition to the eukaryotic genes 30 listed above, the invention, as described, can be used to define molecules that interrupt viral, bacterial or fungal transcription or translation efficiencies and therefore form the basis for a novel anti-infectious disease therapeutic.

In other embodiments, test compounds identified by the methods of the invention are useful for preventing the interaction of an RNA, such as a transfer RNA 35 ("tRNA"), an enzymatic RNA or a ribosomal RNA ("rRNA"), with a protein or with another RNA, thus preventing, e.g., assembly of an *in vivo* protein-RNA or RNA-RNA complex that

is essential for the viability of a cell. The term "enzymatic RNA," as used herein, refers to RNA molecules that are either self-splicing, or that form an enzyme by virtue of their association with one or more proteins, e.g., as in RNase P, telomerase or small nuclear ribonuclear protein particles. For example, inhibition of an interaction between rRNA and one or more ribosomal proteins may inhibit the assembly of ribosomes, rendering a cell incapable of synthesizing proteins. In addition, inhibition of the interaction of precursor rRNA with ribonucleases or ribonucleoprotein complexes (such as RNase P) that process the precursor rRNA prevent maturation of the rRNA and its assembly into ribosomes. Similarly, a tRNA:tRNA synthetase complex may be inhibited by test compounds identified by the methods of the invention such that tRNA molecules do not become charged with amino acids. Such interactions include, but are not limited to, rRNA interactions with ribosomal proteins, tRNA interactions with tRNA synthetase, RNase P protein interactions with RNase P RNA, and telomerase protein interactions with telomerase RNA.

In other embodiments, test compounds identified by the methods of the invention are useful for treating or preventing a viral, bacterial, protozoan or fungal infection. For example, transcriptional up-regulation of the genes of human immunodeficiency virus type 1 ("HIV-1") requires binding of the HIV Tat protein to the HIV trans-activation response region RNA ("TAR RNA"). HIV TAR RNA is a 59-base stem-loop structure located at the 5'-end of all nascent HIV-1 transcripts (Jones & Peterlin, 1994, *Annu. Rev. Biochem.* 63:717-43). Tat protein is known to interact with uracil 23 in the bulge region of the stem of TAR RNA. Thus, TAR RNA is a potential binding target for test compounds, such as small peptides and peptide analogs that bind to the bulge region of TAR RNA and inhibit formation of a Tat-TAR RNA complex involved in HIV-1 upregulation (see Hwang *et al.*, 1999 *Proc. Natl. Acad. Sci. USA* 96:12997-13002). Accordingly, test compounds that bind to TAR RNA are useful as anti-HIV therapeutics (Hamy *et al.*, 1997, *Proc. Natl. Acad. Sci. USA* 94:3548-3553; Hamy *et al.*, 1998, *Biochemistry* 37:5086-5095; Mei *et al.*, 1998, *Biochemistry* 37:14204-14212), and therefore, are useful for treating or preventing AIDS.

The methods of the invention can be used to identify test compounds to treat or prevent viral, bacterial, protozoan or fungal infections in a patient. In some embodiments, the methods of the invention are useful for identifying compounds that decrease translation of microbial genes by interacting with mRNA, as described above, or for identifying compounds that inhibit the interactions of microbial RNAs with proteins or other ligands that are essential for viability of the virus or microbe. Examples of microbial target RNAs useful in the present invention for identifying antiviral, antibacterial, anti-protozoan and anti-fungal compounds include, but are not limited to, general antiviral and anti-inflammatory targets

such as mRNAs of INF α , INF γ , RNase L, RNase L inhibitor protein, PKR, tumor necrosis factor, interleukins 1-15, and IMP dehydrogenase; internal ribosome entry sites; HIV-1 CT rich domain and RNase H mRNA; HCV internal ribosome entry site (required to direct 5 translation of HCV mRNA), and the 3'-untranslated tail of HCV genomes; rotavirus NSP3 binding site, which binds the protein NSP3 that is required for rotavirus mRNA translation; HBV epsilon domain; Dengue virus 5' and 3' untranslated regions, including IRES; INF α , INF β and INF γ ; plasmodium falciparum mRNAs; the 16S ribosomal subunit ribosomal RNA and the RNA component of RNase P of bacteria; and the RNA component of telomerase in 10 fungi and cancer cells. Other target viral and bacterial mRNAs include, but are not limited to, those disclosed in Section 5.

One of skill in the art will appreciate that, although such target RNAs are functionally conserved in various species (*e.g.*, from yeast to humans), they exhibit nucleotide sequence and structural diversity. Therefore, inhibition of, for example, yeast 15 telomerase by an anti-fungal compound identified by the methods of the invention might not interfere with human telomerase and normal human cell proliferation.

Thus, the methods of the invention can be used to identify test compounds that interfere with one or more target RNA interactions with host cell factors that are important for cell growth or viability, or essential in the life cycle of a virus, a bacterium, a 20 protozoa or a fungus. Such test compounds and/or congeners that demonstrate desirable biologic and pharmacologic activity can be administered to a patient in need thereof in order to treat or prevent a disease caused by viral, bacterial, protozoan, or fungal infections. Such diseases include, but are not limited to, HIV infection, AIDS, human T-cell leukemia, SIV infection, FIV infection, feline leukemia, hepatitis A, hepatitis B, hepatitis C, Dengue fever, 25 malaria, rotavirus infection, severe acute gastroenteritis, diarrhea, encephalitis, hemorrhagic fever, syphilis, legionella, whooping cough, gonorrhea, sepsis, influenza, pneumonia, tinea infection, candida infection, and meningitis.

Non-limiting examples of RNA elements involved in the regulation of gene expression, *i.e.*, mRNA stability, translational efficiency via translational initiation and 30 ribosome assembly, *etc.*, include the HIV TAR element, internal ribosome entry site, "slippery site", instability elements, and adenylate uridylate-rich elements, as discussed below.

4.1.1. HIV TAR Element

35 Transcriptional up-regulation of the genes of human immunodeficiency virus type 1 ("HIV-1") requires binding of the HIV Tat protein to the HIV trans-activation

response region RNA (“TAR RNA”), a 59-base stem-loop structure located at the 5' end of all nascent HIV-1 transcripts (Jones & Peterlin, 1994, *Annu. Rev. Biochem.* 63:717-43). Tat protein is known to interact with uracil 23 in the bulge region of the stem of TAR RNA.

5 Thus, TAR RNA is a useful binding target for test compounds, such as small peptides and peptide analogs that bind to the bulge region of TAR RNA and inhibit formation of a Tat-TAR RNA complex involved in HIV-1 up-regulation (see Hwang *et al.*, 1999 *Proc. Natl. Acad. Sci. USA* 96:12997-13002). Accordingly, test compounds that bind to TAR RNA can be useful as anti-HIV therapeutics (Hamy *et al.*, 1997, *Proc. Natl. Acad. Sci. USA* 94:3548-10 3553; Hamy *et al.*, 1998, *Biochemistry* 37:5086-5095; Mei *et al.*, 1998, *Biochemistry* 37:14204-14212), and therefore, are useful for treating or preventing AIDS.

4.1.2. Internal Ribosome Entry Site (“IRES”)

Internal ribosome entry sites (“IRES”) are found in the 5' untranslated regions (“5' UTR”) of several mRNAs, and are thought to be involved in the regulation of 15 translational efficiency. When the IRES element is present on an mRNA downstream of a translational stop codon, it directs ribosomal re-entry (Ghattas *et al.*, 1991, *Mol. Cell. Biol.* 11:5848-5959), which permits initiation of translation at the start of a second open reading frame.

20 As reviewed by Jang *et al.*, a large segment of the 5' nontranslated region, approximately 400 nucleotides in length, promotes internal entry of ribosomes independent of the non-capped 5' end of picornavirus mRNAs (mammalian plus-strand RNA viruses whose genomes serve as mRNA). This 400 nucleotide segment (IRES), maps approximately 200 nt down-stream from the 5' end and is highly structured. IRES elements of different 25 picornaviruses, although functionally similar *in vitro* and *in vivo*, are not identical in sequence or structure. However, IRES elements of the genera entero- and rhinoviruses, on the one hand, and cardio- and aphthoviruses, on the other hand, reveal similarities corresponding to phylogenetic kinship. All IRES elements contain a conserved Yn-Xm-AUG unit (Y, pyrimidine; X, nucleotide) which appears essential for IRES function. 30 The IRES elements of cardio-, entero- and aphthoviruses bind a cellular protein, p57. In the case of cardioviruses, the interaction between a specific stem-loop of the IREs is essential for translation *in vitro*. The IRES elements of entero- and cardioviruses also bind the cellular protein, p52, but the significance of this interaction remains to be shown. The function of p57 or p52 in cellular metabolism is unknown. Since picornaviral IRES elements function *in vivo* 35 in the absence of any viral gene products, is speculated that IRES-like elements may also occur in specific cellular mRNAs releasing them from cap-dependent translation (Jang *et al.*,

1990, Enzyme 44(1-4):292-309).

4.1.3. “Slippery Site”

5 Programmed, or directed, ribosomal frameshifting, when ribosomes shift from one translation reading frame to another and synthesize two viral proteins from a single viral mRNA, is directed by a unique site in viral mRNAs called the “slippery site.” The slippery site directs ribosomal frameshifting in the -1 or +1 direction that causes the ribosome to slip by one base in the 5' direction thereby placing the ribosome in the new reading frame to 10 produce a new protein.

15 Programmed, or directed, ribosomal frameshifting is of particular value to viruses that package their plus strands, as it eliminates the need to splice their mRNAs and reduces the risk of packaging defective genomes and regulates the ratio of viral proteins synthesized. Examples of programmed translational frameshifting (both +1 and -1 shifts) have been identified in ScV systems (Lopinski *et al.*, 2000, Mol. Cell. Biol. 20(4):1095-103, retroviruses (Falk *et al.*, 1993, J. Virol. 67:273-6277; Jacks & Varmus, 1985, Science 230:1237-1242; Morikawa & Bishop, 1992, Virology 186:389-397; Nam *et al.*, 1993, J. Virol. 67:196-203); coronaviruses (Brierley *et al.*, 1987, EMBO J. 6:3779-3785; Herold & Siddell, 1993, Nucleic Acids Res. 21:5838-5842); giardiviruses, which are also members of 20 the Totiviridae (Wang *et al.*, 1993, Proc. Natl. Acad. Sci. USA 90:8595-8599); two bacterial genes (Blinkowa & Walker, 1990, Nucleic Acids Res., 18:1725-1729; Craigen & Caskey, 1986, Nature 322:273); bacteriophage genes (Condron *et al.*, 1991, Nucleic Acids Res. 19:5607-5612); astroviruses (Marczinke *et al.*, 1994, J. Virol. 68:5588-5595); the yeast EST3 gene (Lundblad & Morris, 1997, Curr. Biol. 7:969-976); and the rat, mouse, Xenopus, and 25 Drosophila ornithine decarboxylase antizymes (Matsufuji *et al.*, 1995, Cell 80:51-60); and a significant number of cellular genes (Herold & Siddell, 1993, Nucleic Acids Res. 21:5838-5842).

30 Drugs targeted to ribosomal frameshifting minimize the problem of virus drug resistance because this strategy targets a host cellular process rather than one introduced into the cell by the virus, which minimizes the ability of viruses to evolve drug-resistant mutants. Compounds that target the RNA elements involved in regulating programmed frameshifting should have several advantages, including (a) any selective pressure on the host cellular translational machinery to adapt to the drugs would have to occur at the host evolutionary time scale, which is on the order of millions of years, (b) ribosomal frameshifting is not used 35 to express any host proteins, and (c) altering viral frameshifting efficiencies by modulating

the activity of a host protein minimizing the likelihood that the virus will acquire resistance to such inhibition by mutations in its own genome.

4.1.4. Instability Elements

“Instability elements” may be defined as specific sequence elements that promote the recognition of unstable mRNAs by cellular turnover machinery. Instability elements have been found within mRNA protein coding regions as well as untranslated regions.

10 Altering the control of stability of normal mRNAs may lead to disease. The alteration of mRNA stability has been implicated in diseases such as, but not limited to, cancer, immune disorders, heart disease, and fibrotic disorders.

There are several examples of mutations that delete instability elements which then result in stabilization of mRNAs that may be involved in the onset of cancer. In Burkitt's lymphoma, a portion of the *c-myc* proto-oncogene is translocated to an Ig locus, producing a form of the *c-myc* mRNA that is five times more stable (see, e.g., Kapstein *et al.*, 1996, *J. Biol. Chem.* 271(31):18875-84). The highly oncogenic *v-fos* mRNA lacks the 3' UTR adenylate uridylate rich element ("ARE") that is found in the more labile and weakly oncogenic *c-fos* mRNA (see, e.g., Schiavi *et al.*, 1992, *Biochim Biophys Acta*, 1114(2-3):95-106). Differences between the benign cervical lesions brought about by nonintegrated circular human papillomavirus type 16 and its integrated form, that lacks the 3' UTR ARE and correlates with cervical carcinomas, may be a consequence of stabilizing the E6/E7 transcripts encoding oncogenic proteins. Integration of the virus results in deletion of the ARE instability element, resulting in stabilization of the transcripts and over-expression of the proteins (see, e.g., Jeon & Lambert, 1995, *Proc. Natl. Acad. Sci. USA* 92(5):1654-8). Deletion of AREs from the 3' UTR of the IL-2 and IL-3 genes promotes increased stabilization of these mRNAs, high expression of these proteins, and leads to the formation of cancerous cells (see, e.g., Stoecklin *et al.*, 2000, *Mol. Cell. Biol.* 20(11):3753-63).

Mutations in trans-acting factors involved in mRNA turnover may also 30 promote cancer. In monocytic tumors, the lymphokine GM-CSF mRNA is specifically stabilized as a consequence of an oncogenic lesion in a trans-acting factor that controls mRNA turnover rates. Furthermore, the normally unstable IL-3 transcript is inappropriately long-lived in mast tumor cells. Similarly, the labile GM-CSF mRNA is greatly stabilized in bladder carcinoma cells. See, e.g., Bickel *et al.*, 1990, J. Immunol. 145(3):840-5.

35 The immune system is regulated by a large number of regulatory molecules that either activate or inhibit the immune response. It has now been clearly demonstrated that

stability of the transcripts encoding these proteins are highly regulated. Altered regulation of these molecules leads to mis-regulation of this process and can result in drastic medical consequences. For example, recent results using transgenic mice have shown that mis-
5 regulation of the stability of the important modulator TNF α mRNA leads to diseases such as, but not limited to, rheumatoid arthritis and a Crohn's-like liver disease. *See, e.g.,* Clark, 2000, *Arthritis Res.* 2(3):172-4.

Smooth muscle in the heart is modulated by the β -adrenergic receptor, which in turn responds to the sympathetic neurotransmitter norepinephrine and the adrenal hormone epinephrine. Chronic heart failure is characterized by impairment of smooth muscle cells, which results, in part, from the more rapid decay of the β -adrenergic receptor mRNA. *See, e.g.,* Ellis & Frielle T., 1999, *Biochem. Biophys. Res. Commun.* 258(3):552-8.

10 A large number of diseases result from over-expression of collagen. For example, cirrhosis results from damage to the liver as a consequence of cancer, viral infection, or alcohol abuse. Such damage causes mis-regulation of collagen expression, leading to the formation of large collagen deposits. Recent results indicate that the sizeable increase in collagen expression is largely attributable to stabilization of its mRNA. *See, e.g.,* Lindquist *et al.*, 2000, *Am. J. Physiol. Gastrointest. Liver Physiol.* 279(3):G471-6.

20 4.1.5. Adenylate Uridylate-rich Elements (“ARE”)

Adenylate uridylate-rich elements (“ARE”) are found in the 3' untranslated regions (“3' UTR”) of several mRNAs, and involved in the turnover of mRNAs, such as but not limited to transcription factors, cytokines, and lymphokines. AREs may function both as stabilizing and destabilizing elements. ARE mRNAs are classified into five groups, 25 depending on sequence (Bakheet *et al.*, 2001, *Nucl. Acids Res.* 29(1):246-254). An ongoing database at the web site <http://rc.kfshrc.edu.sa/ared> contains ARE-containing mRNAs and their cluster groups, which is incorporated by reference in its entirety. The ARE motifs are classified as follows:

| | | | |
|----|-------------------|------------------------------|--------------|
| 30 | Group I Cluster | (AUUUAUUUAUUUAUUUAUUA) | SEQ ID NO: 1 |
| | Group II Cluster | (AUUUAUUUAUUUAUUUUA) stretch | SEQ ID NO: 2 |
| | Group III Cluster | (WAUUUAUUUAUUUAW) stretch | SEQ ID NO: 3 |
| | Group IV Cluster | (WWAUUUAUUUAWW) stretch | SEQ ID NO: 4 |
| | Group V Cluster | (WWWWAUUUAWWWWW) stretch | SEQ ID NO: 5 |

35 The ARE-mRNAs were clustered into five groups containing five, four, three and two pentameric repeats, while the last group contains only one pentamer within the

13-bp ARE pattern. Functional categories were assigned whenever possible according to NCBI-COG functional annotation (Tatusov *et al.*, 2001, Nucleic Acids Research, 29(1): 22-28), in addition to the categories: inflammation, immune response, 5 development/differentiation, using an extensive literature search.

Group I contains many secreted proteins including GM-CSF, IL-1, IL-11, IL-12 and Gro- β that affect the growth of hematopoietic and immune cells (Witsell & Schook, 1992, Proc. Natl Acad. Sci. USA, 89:4754-4758). Although TNF α is both a pro-inflammatory and anti-tumor protein, there is experimental evidence that it can act as a 10 growth factor in certain leukemias and lymphomas (Liu *et al.*, 2000, J. Biol. Chem. 275:21086-21093).

Unlike Group I, Groups II-V contain functionally diverse gene families comprising immune response, cell cycle and proliferation, inflammation and coagulation, angiogenesis, metabolism, energy, DNA binding and transcription, nutrient transportation 15 and ionic homeostasis, protein synthesis, cellular biogenesis, signal transduction, and apoptosis (Bakheet *et al.*, 2001, Nucl. Acids Res. 29(1):246-254).

Several groups have described ARE-binding proteins that influence the ARE-mRNA stability. Among the well-characterized proteins are the mammalian homologs of ELAV (embryonic lethal abnormal vision) proteins including AUF1, HuR and He1-N2 20 (Zhang *et al.*, 1993, Mol. Cell. Biol. 13:7652-7665; Levine *et al.*, 1993, Mol. Cell. Biol. 13:3494-3504; Ma *et al.*, 1996, J. Biol. Chem. 271:8144-8151). The zinc-finger protein tristetraprolin has been identified as another ARE-binding protein with destabilizing activity on TNF α , IL-3 and GM-CSF mRNAs (Stoecklin *et al.*, 2000, Mol. Cell. Biol. 20:3753-3763; Carballo *et al.*, 2000, Blood 95:1891-1899).

25 Since ARE-containing genes are clearly important in biological systems, including but not limited to a number of the early response genes that regulate cell proliferation and responses to exogenous agents, the identification of compounds that bind to one or more of the ARE clusters and potentially modulate the stability of the target RNA can potentially be of value as a therapeutic.

30

4.2. Detectably Labeled Target RNAs

Target nucleic acids, including but not limited to RNA and DNA, useful in the methods of the present invention have a label that is detectable via conventional spectroscopic means or radiographic means. Preferably, target nucleic acids are labeled with 35 a covalently attached dye molecule. Useful dye-molecule labels include, but are not limited

to, fluorescent dyes, phosphorescent dyes, ultraviolet dyes, infrared dyes, and visible dyes. Preferably, the dye is a visible dye.

Useful labels in the present invention can include, but are not limited to, spectroscopic labels such as fluorescent dyes (e.g., fluorescein and derivatives such as fluorescein isothiocyanate (FITC) and Oregon Green™, rhodamine and derivatives (e.g., Texas red, tetramethylrhodamine isothiocyanate (TRITC), bora-3a,4a-diaza-s-indacene (BODIPY®) and derivatives, etc.), digoxigenin, biotin, phycoerythrin, AMCA, CyDye™, and the like), radiolabels (e.g., ³H, ¹²⁵I, ³⁵S, ¹⁴C, ³²P, ³³P, etc.), enzymes (e.g., horse radish peroxidase, alkaline phosphatase etc.), spectroscopic colorimetric labels such as colloidal gold or colored glass or plastic (e.g. polystyrene, polypropylene, latex, etc.) beads, or nanoparticles – nanoclusters of inorganic ions with defined dimension from 0.1 to 1000 nm. The label may be coupled directly or indirectly to a component of the detection assay (e.g., the detection reagent) according to methods well known in the art. A wide variety of labels may be used, with the choice of label depending on sensitivity required, ease of conjugation with the compound, stability requirements, available instrumentation, and disposal provisions.

In one embodiment, nucleic acids that are labeled at one or more specific locations are chemically synthesized using phosphoramidite or other solution or solid-phase methods. Detailed descriptions of the chemistry used to form polynucleotides by the phosphoramidite method are well known (see, e.g., Caruthers *et al.*, U.S. Pat. Nos. 4,458,066 and 4,415,732; Caruthers *et al.*, 1982, *Genetic Engineering* 4:1-17; *Users Manual Model 392 and 394 Polynucleotide Synthesizers*, 1990, pages 6-1 through 6-22, Applied Biosystems, Part No. 901237; Ojwang, *et al.*, 1997, *Biochemistry*, 36:6033-6045). The phosphoramidite method of polynucleotide synthesis is the preferred method because of its efficient and rapid coupling and the stability of the starting materials. The synthesis is performed with the growing polynucleotide chain attached to a solid support, such that excess reagents, which are generally in the liquid phase, can be easily removed by washing, decanting, and/or filtration, thereby eliminating the need for purification steps between synthesis cycles.

The following briefly describes illustrative steps of a typical polynucleotide synthesis cycle using the phosphoramidite method. First, a solid support to which is attached a protected nucleoside monomer at its 3' terminus is treated with acid, e.g., trichloroacetic acid, to remove the 5'-hydroxyl protecting group, freeing the hydroxyl group for a subsequent coupling reaction. After the coupling reaction is completed an activated intermediate is formed by contacting the support-bound nucleoside with a protected nucleoside phosphoramidite monomer and a weak acid, e.g., tetrazole. The weak acid

protonates the nitrogen atom of the phosphoramidite forming a reactive intermediate. Nucleoside addition is generally complete within 30 seconds. Next, a capping step is performed, which terminates any polynucleotide chains that did not undergo nucleoside addition. Capping is preferably performed using acetic anhydride and 1-methylimidazole. The phosphite group of the internucleotide linkage is then converted to the more stable phosphotriester by oxidation using iodine as the preferred oxidizing agent and water as the oxygen donor. After oxidation, the hydroxyl protecting group of the newly added nucleoside is removed with a protic acid, *e.g.*, trichloroacetic acid or dichloroacetic acid, and the cycle is repeated one or more times until chain elongation is complete. After synthesis, the polynucleotide chain is cleaved from the support using a base, *e.g.*, ammonium hydroxide or *t*-butyl amine. The cleavage reaction also removes any phosphate protecting groups, *e.g.*, cyanoethyl. Finally, the protecting groups on the exocyclic amines of the bases and any protecting groups on the dyes are removed by treating the polynucleotide solution in base at an elevated temperature, *e.g.*, at about 55°C. Preferably the various protecting groups are removed using ammonium hydroxide or *t*-butyl amine.

Any of the nucleoside phosphoramidite monomers can be labeled using standard phosphoramidite chemistry methods (Hwang *et al.*, 1999, Proc. Natl. Acad. Sci. USA 96(23):12997-13002; Ojwang *et al.*, 1997, Biochemistry. 36:6033-6045 and references cited therein). Dye molecules useful for covalently coupling to phosphoramidites preferably comprise a primary hydroxyl group that is not part of the dye's chromophore. Illustrative dye molecules include, but are not limited to, disperse dye CAS 4439-31-0, disperse dye CAS 6054-58-6, disperse dye CAS 4392-69-2 (Sigma-Aldrich, St. Louis, MO), disperse red, and 1-pyrenebutanol (Molecular Probes, Eugene, OR). Other dyes useful for coupling to phosphoramidites will be apparent to those of skill in the art, such as fluorescein, cy3, and cy5 fluorescent dyes, and may be purchased from, *e.g.*, Sigma-Aldrich, St. Louis, MO or Molecular Probes, Inc., Eugene, OR.

In another embodiment, dye-labeled target RNA molecules are synthesized enzymatically using *in vitro* transcription (Hwang *et al.*, 1999, Proc. Natl. Acad. Sci. USA 96(23):12997-13002 and references cited therein). In this embodiment, a template DNA is denatured by heating to about 90°C and an oligonucleotide primer is annealed to the template DNA, for example by slow-cooling the mixture of the denatured template and the primer from about 90°C to room temperature. A mixture of ribonucleoside-5'-triphosphates capable of supporting template-directed enzymatic extension of the primed template (*e.g.*, a mixture including GTP, ATP, CTP, and UTP), including one or more dye-labeled ribonucleotides (Sigma-Aldrich, St. Louis, MO), is added to the primed template. Next, a polymerase

enzyme is added to the mixture under conditions where the polymerase enzyme is active, which are well-known to those skilled in the art. A labeled polynucleotide is formed by the incorporation of the labeled ribonucleotides during polymerase-mediated strand synthesis.

5 In yet another embodiment of the invention, nucleic acid molecules are end-labeled after their synthesis. Methods for labeling the 5'-end of an oligonucleotide include but are by no means limited to: (i) periodate oxidation of a 5'-to-5'-coupled ribonucleotide, followed by reaction with an amine-reactive label (Heller & Morisson, 1985, in *Rapid Detection and Identification of Infectious Agents*, D.T. Kingsbury and S. Falkow, eds., pp. 245-256, Academic Press); (ii) condensation of ethylenediamine with 5'-phosphorylated 10 polynucleotide, followed by reaction with an amine reactive label (Morrison, European Patent Application 232 967); (iii) introduction of an aliphatic amine substituent using an aminoethyl phosphite reagent in solid-phase DNA synthesis, followed by reaction with an amine reactive label (Cardullo *et al.*, 1988, Proc. Natl. Acad. Sci. USA 85:8790-8794); and 15 (iv) introduction of a thiophosphate group on the 5'-end of the nucleic acid, using phosphatase treatment followed by end-labeling with ATP- S and kinase, which reacts specifically and efficiently with maleimide-labeled fluorescent dyes (Czworkowski *et al.*, 1991, Biochem. 30:4821-4830).

20 A detectable label should not be incorporated into a target nucleic acid at the specific binding site at which test compounds are likely to bind, since the presence of a 25 covalently attached label might interfere sterically or chemically with the binding of the test compounds at this site. Accordingly, if the region of the target nucleic acid that binds to a host cell factor is known, a detectable label is preferably incorporated into the nucleic acid molecule at one or more positions that are spatially or sequentially remote from the binding region.

30 After synthesis, the labeled target nucleic acid can be purified using standard techniques known to those skilled in the art (see Hwang *et al.*, 1999, Proc. Natl. Acad. Sci. USA 96(23):12997-13002 and references cited therein). Depending on the length of the target nucleic acid and the method of its synthesis, such purification techniques include, but are not limited to, reverse-phase high-performance liquid chromatography ("reverse-phase HPLC"), fast performance liquid chromatography ("FPLC"), and gel purification. After purification, the target RNA is refolded into its native conformation, preferably by heating to approximately 85-95°C and slowly cooling to room temperature in a buffer, e.g., a buffer comprising about 50 mM Tris-HCl, pH 8 and 100 mM NaCl.

35 In another embodiment, the target nucleic acid can also be radiolabeled. A radiolabel, such as, but not limited to, an isotope of phosphorus, sulfur, or hydrogen, may be

incorporated into a nucleotide, which is added either after or during the synthesis of the target nucleic acid. Methods for the synthesis and purification of radiolabeled nucleic acids are well known to one of skill in the art. See, e.g., Sambrook *et al.*, 1989, in *Molecular Cloning: A Laboratory Manual*, pp 10.2-10.70, Cold Spring Harbor Laboratory Press, and the references cited therein, which are hereby incorporated by reference in their entireties.

5 In another embodiment, the target nucleic acid can be attached to an inorganic nanoparticle. A nanoparticle is a cluster of ions with controlled size from 0.1 to 1000 nm comprised of metals, metal oxides, or semiconductors including, but not limited to Ag_2S ,
10 ZnS , CdS , CdTe , Au , or TiO_2 . Nanoparticles have unique optical, electronic and catalytic properties relative to bulk materials which can be adjusted according to the size of the particle. Methods for the attachment of nucleic acids are well known to one of skill in the art (see, e.g., Niemeyer, 2001, *Angew. Chem. Int. Ed.* 40: 4129-4158, International Patent Publication WO/0218643, and the references cited therein, the disclosures of which are hereby incorporated by reference in their entireties).

15

4.3. Libraries of Small Molecules

Libraries screened using the methods of the present invention can comprise a variety of types of test compounds on solid supports. In all of the embodiments described 20 below, all of the libraries can be synthesized on solid supports or the compounds of the library can be attached to solid supports by linkers.

In some embodiments, the test compounds are nucleic acid or peptide molecules. In a non-limiting example, peptide molecules can exist in a phage display library. In other embodiments, types of test compounds include, but are not limited to, peptide 25 analogs including peptides comprising non-naturally occurring amino acids, e.g., D-amino acids, phosphorous analogs of amino acids, such as α -amino phosphoric acids and α -amino phosphoric acids, or amino acids having non-peptide linkages, nucleic acid analogs such as phosphorothioates and PNAs, hormones, antigens, synthetic or naturally occurring drugs, opiates, dopamine, serotonin, catecholamines, thrombin, acetylcholine, prostaglandins, 30 organic molecules, pheromones, adenosine, sucrose, glucose, lactose and galactose. Libraries of polypeptides or proteins can also be used.

In a preferred embodiment, the combinatorial libraries are small organic molecule libraries, such as, but not limited to, benzodiazepines, isoprenoids, thiazolidinones, metathiazanones, pyrrolidines, morpholino compounds, and diazepindiones. In another 35 embodiment, the combinatorial libraries comprise peptoids; random bio-oligomers; benzodiazepines; diversomers such as hydantoins, benzodiazepines and dipeptides;

vinylogous polypeptides; nonpeptidal peptidomimetics; oligocarbamates; peptidyl phosphonates; peptide nucleic acid libraries; antibody libraries; or carbohydrate libraries. Combinatorial libraries are themselves commercially available (see, e.g., Advanced 5 ChemTech Europe Ltd., Cambridgeshire, UK; ASINEX, Moscow Russia; BioFocus plc, Sittingbourne, UK; Bionet Research (A division of Key Organics Limited), Camelford, UK; ChemBridge Corporation, San Diego, California; ChemDiv Inc, San Diego, California.; ChemRx Advanced Technologies, South San Francisco, California; ComGenex Inc., Budapest, Hungary; Evotec OAI Ltd, Abingdon, UK; IF LAB Ltd., Kiev, Ukraine; 10 Maybridge plc, Cornwall, UK; PharmaCore, Inc., North Carolina; SIDDCO Inc, Tucson, Arizona; TimTec Inc, Newark, Delaware; Tripos Receptor Research Ltd, Bude, UK; Toslab, Ekaterinburg, Russia).

In one embodiment, the combinatorial compound library for the methods of the present invention may be synthesized. There is a great interest in synthetic methods 15 directed toward the creation of large collections of small organic compounds, or libraries, which could be screened for pharmacological, biological or other activity (Dolle, 2001, *J. Comb. Chem.* 3:477-517; Hall *et al.*, 2001, *ibid.* 3:125-150; Dolle, 2000, *ibid.* 2:383-433; Dolle, 1999, *ibid.* 1:235-282). The synthetic methods applied to create vast combinatorial 20 libraries are performed in solution or in the solid phase, *i.e.*, on a solid support. Solid-phase synthesis makes it easier to conduct multi-step reactions and to drive reactions to completion with high yields because excess reagents can be easily added and washed away after each 25 reaction step. Solid-phase combinatorial synthesis also tends to improve isolation, purification and screening. However, the more traditional solution phase chemistry supports a wider variety of organic reactions than solid-phase chemistry. Methods and strategies for the synthesis of combinatorial libraries can be found in *A Practical Guide to Combinatorial Chemistry*, A. W. Czarnik and S.H. Dewitt, eds., American Chemical Society, 1997; *The Combinatorial Index*, B.A. Bunin, Academic Press, 1998; *Organic Synthesis on Solid Phase*, F.Z. Dörwald, Wiley-VCH, 2000; and *Solid-Phase Organic Syntheses*, Vol. 1, A.W. Czarnik, ed., Wiley Interscience, 2001.

30 Combinatorial compound libraries of the present invention may be synthesized using apparatuses described in US Patent No. 6,358,479 to Frisina *et al.*, U.S. Patent No. 6,190,619 to Kilcoin *et al.*, US Patent No. 6,132,686 to Gallup *et al.*, US Patent No. 6,126,904 to Zuellig *et al.*, US Patent No. 6,074,613 to Harness *et al.*, US Patent No. 6,054,100 to Stanchfield *et al.*, and US Patent No. 5,746,982 to Saneii *et al.* which are hereby 35 incorporated by reference in their entirety. These patents describe synthesis apparatuses

capable of holding a plurality of reaction vessels for parallel synthesis of multiple discrete compounds or for combinatorial libraries of compounds.

In one embodiment, the combinatorial compound library can be synthesized in solution. The method disclosed in U.S. Patent No. 6,194,612 to Boger *et al.*, which is hereby incorporated by reference in its entirety, features compounds useful as templates for solution phase synthesis of combinatorial libraries. The template is designed to permit reaction products to be easily purified from unreacted reactants using liquid/liquid or solid/liquid extractions. The compounds produced by combinatorial synthesis using the template will preferably be small organic molecules. Some compounds in the library may mimic the effects of non-peptides or peptides. In contrast to solid phase synthesis of combinatorial compound libraries, liquid phase synthesis does not require the use of specialized protocols for monitoring the individual steps of a multistep solid phase synthesis (Egner *et al.*, 1995, *J. Org. Chem.* 60:2652; Anderson *et al.*, 1995, *J. Org. Chem.* 60:2650; Fitch *et al.*, 1994, *J. Org. Chem.* 59:7955; Look *et al.*, 1994, *J. Org. Chem.* 49:7588; Metzger *et al.*, 1993, *Angew. Chem., Int. Ed. Engl.* 32:894; Youngquist *et al.*, 1994, *Rapid Commun. Mass Spect.* 8:77; Chu *et al.*, 1995, *J. Am. Chem. Soc.* 117:5419; Brummel *et al.*, 1994, *Science* 264:399; Stevanovic *et al.*, 1993, *Bioorg. Med. Chem. Lett.* 3:431).

Combinatorial compound libraries useful for the methods of the present invention can be synthesized on solid supports. In one embodiment, a split synthesis method, a protocol of separating and mixing solid supports during the synthesis, is used to synthesize a library of compounds on solid supports (see Lam *et al.*, 1997, *Chem. Rev.* 97:41-448; Ohlmeyer *et al.*, 1993, *Proc. Natl. Acad. Sci. USA* 90:10922-10926 and references cited therein). Each solid support in the final library has substantially one type of test compound attached to its surface. Other methods for synthesizing combinatorial libraries on solid supports, wherein one product is attached to each support, will be known to those of skill in the art (see, e.g., Nefzi *et al.*, 1997, *Chem. Rev.* 97:449-472 and US Patent No. 6,087,186 to Cargill *et al.* which are hereby incorporated by reference in their entirety).

As used herein, the term "solid support" is not limited to a specific type of solid support. Rather a large number of supports are available and are known to one skilled in the art. Solid supports include silica gels, resins, derivatized plastic films, glass beads, cotton, plastic beads, polystyrene beads, doped polystyrene beads (as described by Fenniri *et al.*, 2000, *J. Am. Chem. Soc.* 123:8151-8152), alumina gels, and polysaccharides. A suitable solid support may be selected on the basis of desired end use and suitability for various synthetic protocols. For example, for peptide synthesis, a solid support can be a resin such as p-methylbenzhydrylamine (pMBHA) resin (Peptides International, Louisville, KY),

polystyrenes (e.g., PAM-resin obtained from Bachem Inc., Peninsula Laboratories, etc.), including chloromethylpolystyrene, hydroxymethylpolystyrene and aminomethylpolystyrene, poly (dimethylacrylamide)-grafted styrene co-divinyl-benzene (e.g., POLYHIPE resin, obtained from Aminotech, Canada), polyamide resin (obtained from Peninsula Laboratories), polystyrene resin grafted with polyethylene glycol (e.g., TENTAGEL or ARGOGEL, Bayer, Tubingen, Germany) polydimethylacrylamide resin (obtained from Milligen/Bioscience, California), or Sepharose (Pharmacia, Sweden). In another embodiment, the solid support can be a magnetic bead coated with streptavidin, such as Dynabeads Streptavidin (Dynal Biotech, Oslo, Norway).

In one embodiment, the solid phase support is suitable for *in vivo* use, *i.e.*, it can serve as a carrier or support for administration of the test compound to a patient (e.g., TENTAGEL, Bayer, Tubingen, Germany). In a particular embodiment, the solid support is palatable and/or orally ingestable.

In some embodiments of the present invention, compounds can be attached to solid supports via linkers. Linkers can be integral and part of the solid support, or they may be nonintegral that are either synthesized on the solid support or attached thereto after synthesis. Linkers are useful not only for providing points of test compound attachment to the solid support, but also for allowing different groups of molecules to be cleaved from the solid support under different conditions, depending on the nature of the linker. For example, linkers can be, *inter alia*, electrophilically cleaved, nucleophilically cleaved, photocleavable, enzymatically cleaved, cleaved by metals, cleaved under reductive conditions or cleaved under oxidative conditions.

4.4. Library Screening

After a target nucleic acid, such as but not limited to RNA or DNA, is labeled and a test compound library is synthesized or purchased or both, the labeled target nucleic acid is used to screen the library to identify test compounds that bind to the nucleic acid. Screening comprises contacting a labeled target nucleic acid with an individual, or small group, of the components of the compound library. Preferably, the contacting occurs in an aqueous solution, and most preferably, under physiologic conditions. The aqueous solution preferably stabilizes the labeled target nucleic acid and prevents denaturation or degradation of the nucleic acid without interfering with binding of the test compounds. The aqueous solution can be similar to the solution in which a complex between the target RNA and its corresponding host cell factor is formed *in vitro*. For example, TK buffer, which is commonly used to form Tat protein-TAR RNA complexes *in vitro*, can be used in the

methods of the invention as an aqueous solution to screen a library of test compounds for TAR RNA binding compounds.

The methods of the present invention for screening a library of test compounds preferably comprise contacting a test compound with a target nucleic acid in the presence of an aqueous solution, the aqueous solution comprising a buffer and a combination of salts, preferably approximating or mimicking physiologic conditions. The aqueous solution optionally further comprises non-specific nucleic acids, such as, but not limited to, DNA; yeast tRNA; salmon sperm DNA; homoribopolymers such as, but not limited to, poly IC, polyA, polyU, and polyC; and non-specific RNA. The non-specific RNA may be an unlabeled target nucleic acid having a mutation at the binding site, which renders the unlabeled nucleic acid incapable of interacting with a test compound at that site. For example, if dye-labeled TAR RNA is used to screen a library, unlabeled TAR RNA having a mutation in the uracil 23/cytosine 24 bulge region may also be present in the aqueous solution. Without being bound by any theory, the addition of unlabeled RNA that is essentially identical to the dye-labeled target RNA except for a mutation at the binding site might minimize interactions of other regions of the dye-labeled target RNA with test compounds or with the solid support and prevent false positive results.

The solution further comprises a buffer, a combination of salts, and optionally, a detergent or a surfactant. The pH of the solution typically ranges from about 5 to about 8, preferably from about 6 to about 8, most preferably from about 6.5 to about 8. A variety of buffers may be used to achieve the desired pH. Suitable buffers include, but are not limited to, Tris, Mes, Bis-Tris, Ada, Aces, Pipes, Mopso, Bis-Tris propane, Bes, Mops, Tes, Hepes, Dipso, Mobs, Tapso, Trizma, Heppso, Popso, TEA, Epps, Tricine, Gly-Gly, Bicine, and sodium-potassium phosphate. The buffering agent comprises from about 10 mM to about 100 mM, preferably from about 25 mM to about 75 mM, most preferably from about 40 mM to about 60 mM buffering agent. The pH of the aqueous solution can be optimized for different screening reactions, depending on the target RNA used and the types of test compounds in the library, and therefore, the type and amount of the buffer used in the solution can vary from screen to screen. In a preferred embodiment, the aqueous solution has a pH of about 7.4, which can be achieved using about 50 mM Tris buffer.

In addition to an appropriate buffer, the aqueous solution further comprises a combination of salts, from about 0 mM to about 100 mM KCl, from about 0 mM to about 1 M NaCl, and from about 0 mM to about 200 mM MgCl₂. In a preferred embodiment, the combination of salts is about 100 mM KCl, 500 mM NaCl, and 10 mM MgCl₂. Without being bound by any theory, Applicant has found that a combination of KCl, NaCl, and MgCl₂

stabilizes the target RNA such that most of the RNA is not denatured or digested over the course of the screening reaction. The optional concentration of each salt used in the aqueous solution is dependent on the particular target RNA used and can be determined using routine experimentation.

5 The solution optionally comprises from about 0.01% to about 0.5% (w/v) of a detergent or a surfactant. Without being bound by any theory, a small amount of detergent or surfactant in the solution might reduce non-specific binding of the target RNA to the solid support and control aggregation and increase stability of target RNA molecules. Typical detergents useful in the methods of the present invention include, but are not limited to, anionic detergents, such as salts of deoxycholic acid, 1-heptanesulfonic acid, N-laurylsarcosine, lauryl sulfate, 1-octane sulfonic acid and taurocholic acid; cationic detergents such as benzalkonium chloride, cetylpyridinium, methylbenzethonium chloride, and decamethonium bromide; zwitterionic detergents such as CHAPS, CHAPSO, alkyl betaines, alkyl amidoalkyl betaines, N-dodecyl-N,N-dimethyl-3-ammonio-1-propanesulfonate, and phosphatidylcholine; and non-ionic detergents such as n-decyl α -D-glucopyranoside, n-decyl β -D-maltopyranoside, n-dodecyl β -D-maltoside, n-octyl β -D-glucopyranoside, sorbitan esters, n-tetradecyl β -D-maltoside, octylphenoxy polyethoxyethanol (Nonidet P-40), nonylphenoxy polyethoxyethanol (NP-40), and tritons. Preferably, the detergent, if present, is a nonionic detergent. Typical surfactants useful in the methods of the present invention include, but are not limited to, ammonium lauryl sulfate, polyethylene glycols, butyl glucoside, decyl glucoside, Polysorbate 80, lauric acid, myristic acid, palmitic acid, potassium palmitate, undecanoic acid, lauryl betaine, and lauryl alcohol. More preferably, the detergent, if present, is Triton X-100 and present in an amount of about 0.1% (w/v).

25 Non-specific binding of a labeled target nucleic acid to test compounds can be further minimized by treating the binding reaction with one or more blocking agents. In one embodiment, the binding reactions are treated with a blocking agent, *e.g.*, bovine serum albumin ("BSA"), before contacting with the labeled target nucleic acid. In another embodiment, the binding reactions are treated sequentially with at least two different blocking agents. This blocking step is preferably performed at room temperature for from about 0.5 to about 3 hours. In a subsequent step, the reaction mixture is further treated with unlabeled RNA having a mutation at the binding site. This blocking step is preferably performed at about 4°C for from about 12 hours to about 36 hours before addition of the dye-labeled target RNA. Preferably, the solution used in the one or more blocking steps is substantially similar to the aqueous solution used to screen the library with the dye-labeled target RNA, *e.g.*, in pH and salt concentration.

Once contacted, the mixture of labeled target nucleic acid and the test compound is preferably maintained at 4°C for from about 1 day to about 5 days, preferably from about 2 days to about 3 days with constant agitation. To identify the reactions in which binding to the labeled target nucleic acid occurred, after the incubation period, bound from free compounds are determined using any of the methods disclosed in Section 4.5 *infra*.
5

4.5. Separation Methods for Screening Test Compounds

After the labeled target RNA is contacted with the library of test compounds immobilized on beads, the beads must then be separated from the unbound target RNA in the liquid phase. This can be accomplished by any number of physical means; *e.g.*, sedimentation, centrifugation. Thereafter, a number of methods can be used to separate the library beads that are complexed with the labeled target RNA from uncomplexed beads in order to isolate the test compound on the bead. Alternatively, mass spectroscopy and NMR spectroscopy can be used to simultaneously identify and separate beads complexed to the labeled target RNA from uncomplexed beads.
10
15

4.5.1. Flow Cytometry

In a preferred embodiment, the complexed and non-complexed target nucleic acids are separated by flow cytometry methods. Flow cytometers for sorting and examining biological cells are well known in the art; this technology can be applied to separate the labeled library beads from unlabeled beads. Known flow cytometers are described, for example, in U.S. Patent Nos. 4,347,935; 5,464,581; 5,483,469; 5,602,039; 5,643,796; and 6,211,477; the entire contents of which are incorporated by reference herein. Other known flow cytometers are the FACS Vantage™ system manufactured by Becton Dickinson and Company, and the COPAS™ system manufactured by Union Biometrica.
20
25

A flow cytometer typically includes a sample reservoir for receiving a biological sample. The biological sample contains particles (hereinafter referred to as "beads") that are to be analyzed and sorted by the flow cytometer. Beads are transported from the sample reservoir at high speed (>100beads/second) to a flow cell in a stream of liquid "sheath" fluid. High-frequency vibrations of a nozzle that directs the stream to the flow cell causes the stream to partition and form ordered droplets, with each droplet containing a single bead. Physical properties of beads can be measured as they intersect a laser beam within the cytometer flow cell. As beads move one by one through the interrogation point, they cause the laser light to scatter and fluorescent molecules on the labeled beads (*i.e.*, beads complexed with labeled target RNA) become excited.
30
35

Alternatively, if the target nucleic acid is labeled with an inorganic nanoparticle, the beads complexed with bound target nucleic acid can be distinguished not only by unique fluorescent properties but also on the basis of spectrometric properties (e.g. including but not limited to increased optical density due to the reduction of Ag^+ ions in the presence of gold nanoparticles (see, e.g., Taton *et al.* *Science* 2000, 289: 1757-1760)).

An appropriate detection system consisting of photomultiplier tubes, photodiodes or other devices for measuring light are focused onto the interrogation point where the properties are measured. In so doing, information regarding particle size (light scatter) and complex formation (fluorescence intensity) is obtained. Particles with the desired physical properties are then sorted by a variety of physical means. In one embodiment, the beads are sorted by an electrostatic method. To sort beads by an electrostatic method, the droplets containing the beads with the desired physical properties are electrically charged and deflected from the trajectory of uncharged droplets as they pass through an electrostatic field formed by two deflection plates held constant at a high electrical potential difference. In another embodiment, the beads are sorted by an air-diverting method. To sort beads by an air-diverting method, the droplets containing the beads with the desired physical properties are deflected from their trajectory by a focused stream of forced air. Both of these embodiments cause the trajectory of beads with the desired physical properties to become changed, thereby sorting them from other beads. Accordingly, the beads complexed to the labeled target RNA can be collected in an appropriate collecting vessel.

Thus, in one embodiment of the present invention, the complexed and non-complexed target nucleic acids are separated by flow cytometry methods. In a preferred embodiment, the target nucleic acid is labeled with a fluorescent label and the complexed and non-complexed target nucleic acids are separated by fluorescence activated cell sorting ("FACS"). Such methods are well known to one of skill in the art.

4.5.2. Affinity Chromatography

In another embodiment of the invention, the target RNA can be labeled with biotin, an antigen, or a ligand. Library beads complexed to the target RNA can be separated from uncomplexed beads using affinity techniques designed to capture the labeled moiety on the target RNA. For example, a solid support, such as but not limited to, a column or a well in a microwell plate coated with avidin/streptavidin, an antibody to the antigen, or a receptor for the ligand can be used to capture or immobilize the labeled beads. Complexed RNA may or may not be irreversibly bound to the bead by a further transformation between the bound

RNA and an additional moiety on the surface of the bead. Such linking methods include, but are not limited to: photochemical crosslinking between RNA and bead-bound molecules such as psoralen, thymidine or uridine derivates either present as monomers, oligomers, or as 5 a partially complementary sequence; or chemical ligation by disulfide exchange, nitrogen mustards, bond formation between an electrophile and a nucleophile, or alkylating reagents. See, *e.g.*, International Patent Publication WO/0146461, the contents of which are hereby incorporated by reference. The unbound library beads can be removed after the binding 10 reaction by washing the solid phase. If the RNA is irreversibly bound to the bead, test compounds can be isolated from the bead following destruction of the bound RNA by preferably, but not limited to, enzymatic or chemical (*e.g.*, alkaline hydrolysis) degradation. The library beads bound to the solid phase can then be eluted with any solution that disrupts 15 the binding between the labeled target RNA and the solid phase. Such solutions include high salt solutions, low pH solutions, detergents, and chaotropic denaturants, and are well known to one of skill in the art. In another embodiment, the test compounds can be eluted from the solid phase by heat.

In one embodiment, the library of test compounds can be prepared on magnetic beads, such as Dynabeads Streptavidin (Dynal Biotech, Oslo, Norway). The magnetic bead library can then be mixed with the labeled target RNA under conditions that 20 allow binding to occur. The separation of the beads from unbound target RNA in the liquid phase can be accomplished using a magnet. After removal of the magnetic field, the bead complexed to the labeled RNA may be separated from uncomplexed library beads via the label used on the target RNA; *e.g.*, biotinylated target RNA can be captured by avidin/streptavidin; target RNA labeled with antigen can be captured by the appropriate 25 antibody; target RNA labeled with ligand can be captured using the appropriate immobilized receptor. The captured library bead can then be eluted with any solution that disrupts the binding between the labeled target RNA and the immobilized surface. Such solutions include high salt solutions, low pH solutions, detergents, and chaotropic denaturants, and are well known to one of skill in the art. Complexed RNA may or may not be irreversibly bound 30 to the bead by a further transformation between the bound RNA and an additional moiety on the surface of the bead. Such linking methods include, but are not limited to: photochemical crosslinking between RNA and bead-bound molecules such as psoralen, thymidine or uridine derivates either present as monomers, oligomers, or as a partially complementary sequence; or chemical ligation by disulfide exchange, nitrogen mustards, bond formation between an 35 electrophile and a nucleophile, or alkylating reagents. See, *e.g.*, International Patent Publication WO/0146461, the contents of which are hereby incorporated by reference. If the

RNA is irreversibly bound to the bead, test compounds can be isolated from the bead following destruction of the bound RNA by enzymatic degradation including, but not limited to, ribonucleases A, U₂, CL₃, T₁, Phy M, *B. cereus* or chemical degradation including, but not limited to, piperidine-promoted backbone cleavage of abasic sites (following treatment with sodium hydroxide, hydrazine, piperidine formate, or dimethyl sulfate), or metal-assisted (e.g. nickel(II), cobalt(II), or iron(II)) oxidative cleavage.

In another embodiment, the preselected target RNA can be labeled with a heavy metal tag and incubated with the library beads to allow binding of the test compounds to the target RNA. The separation of the labeled beads from unlabeled beads can be accomplished using a magnetic field. After removal of the magnetic field, the test compound can be eluted with any solution that disrupts the binding between the preselected target RNA and the test compound. Such solutions include high salt solutions, low pH solutions, detergents, and chaotropic denaturants, and are well known to one of skill in the art. In another embodiment, the test compounds can be eluted from the solid phase by heat.

4.5.3. Manual Batch

In one embodiment, a manual "batch" mode is used for separating complexed beads. To explore a bead-based library within a reasonable time period, the primary screens should be operated with sufficient throughput. To do this, the target nucleic acid is labeled with a dye and then incubated with the combinatorial library. An advantage of such an assay is the fast identification of active library beads by color change. In the lower concentrations of the dye-labeled target molecule, only those library beads that bind the target molecules most tightly are detected because of higher local concentration of the dye. When washed and plated into a liquid monolayer, colored beads are easily separated from non-colored beads with the aid of a dissecting microscope. One of the problems associated with this method could be the interaction between the red dye and library substrates. Control experiments using the dye alone and dye attached to mutant RNA sequences with the libraries are performed to eliminate this possibility.

30

4.5.4. Suspension of Beads in Electric Fields

In another embodiment of the invention, library beads bound to the target RNA can be separated from unbound beads on the basis of the altered charge properties due to RNA binding. In a preferred embodiment of this technique, beads are separated from unbound nucleic acid and suspended, preferably but not only, in the presence of an electric field where the bound RNA causes the beads bound to the target RNA to migrate toward the

anode, or positive, end of the field.

Beads can be preferentially suspended in solution as a colloidal suspension with the aid of detergents or surfactants. Typical detergents useful in the methods of the present invention include, but are not limited to, anionic detergents, such as salts of deoxycholic acid, 1-heptanesulfonic acid, N-laurylsarcosine, lauryl sulfate, 1-octane sulfonic acid, carboxymethylcellulose, carrageenan, and taurocholic acid; cationic detergents such as benzalkonium chloride, cetylpyridinium, methylbenzethonium chloride, and decamethonium bromide; zwitterionic detergents such as CHAPS, CHAPSO, alkyl betaines, alky amidoalkyl betaines, N-dodecyl-N,N-dimethyl-3-ammonio-1-propanesulfonate, and phosphatidylcholine; and non-ionic detergents such as n-decyl α -D-glucopyranoside, n-decyl-D-maltopyranoside, n-dodecyl -D-maltoside, n-octyl -D-glucopyranoside, sorbitan esters, n-tetradecyl -D-maltoside and tritons. Preferably, the detergent, if present, is a nonionic detergent. Typical surfactants useful in the methods of the present invention include, but are not limited to, ammonium lauryl sulfate, polyethylene glycols, butyl glucoside, decyl glucoside, Polysorbate 80, lauric acid, myristic acid, palmitic acid, potassium palmitate, undecanoic acid, lauryl betaine, and lauryl alcohol.

Complexed RNA may or may not be irreversibly bound to the bead by a further transformation between the bound RNA and an additional moiety on the surface of the bead. Such linking methods include, but are not limited to: photochemical crosslinking between RNA and bead-bound molecules such as psoralen, thymidine or uridine derivates either present as monomers, oligomers, or as a partially complementary sequence; or chemical ligation by disulfide exchange, nitrogen mustards, bond formation between an electrophile and a nucleophile, or alkylating reagents.

If the RNA is irreversibly bound to the bead, test compounds can be isolated from the bead following destruction of the bound RNA by enzymatic degradation including, but not limited to, ribonucleases A, U₂, CL₃, T₁, Phy M, *B. cereus* or chemical degradation including, but not limited to, piperidine-promoted backbone cleavage of abasic sites (following treatment with sodium hydroxide, hydrazine, piperidine formate, or dimethyl sulfate), or metal-assisted (e.g. nickel(II), cobalt(II), or iron(II)) oxidative cleavage.

4.5.5. Microwave

In another embodiment, the complexed beads are separated from uncomplexed beads by microwave. For example, as described in U.S. Patent Nos. 6,340,568; 6,338,968; and 6,287,874 to Hefti, the disclosures of which are hereby incorporated by reference, a system which is sensitive to the unique dielectric properties of

molecules and binding complexes, such as hybridization complexes formed between a nucleic acid probe and a nucleic acid target, molecular binding events, and protein/ligand complexes, can be used to analyze nucleic acids. In this system, the different hybridization complexes can be directly distinguished without the use of labels. The method involves 5 contacting a nucleic acid probe that is electromagnetically coupled to a portion of a signal path with a sample containing a target nucleic acid. The portion of the signal path to which the nucleic acid probe is coupled typically is a continuous transmission line. A response signal is detected for a hybridization complex formed between the nucleic acid probe and the 10 nucleic acid target. Detection may involve propagating a test signal along the signal path and then detecting a response signal formed through modulation of the test signal by the hybridization complex.

4.6. Methods for Identifying Test Compounds

15 If the library is a peptide or nucleic acid library, the sequence of the test compound on the isolated bead can be determined by direct sequencing of the peptide or nucleic acid. Such methods are well known to one of skill in the art.

4.6.1. Mass Spectrometry

20 Mass spectrometry (e.g., electrospray ionization ("ESI") and matrix-assisted laser desorption-ionization ("MALDI"), Fourier-transform ion cyclotron resonance ("FT-ICR")) can be used both for high-throughput screening of test compounds that bind to a target RNA and elucidating the structure of the test compound on the isolated bead.

25 MALDI uses a pulsed laser for desorption of the ions and a time-of-flight analyzer, and has been used for the detection of noncovalent tRNA:amino-acyl-tRNA synthetase complexes (Gruic-Sovulj *et al.*, 1997, *J. Biol. Chem.* 272:32084-32091). However, covalent cross-linking between the target nucleic acid and the test compound is required for detection, since a non-covalently bound complex may dissociate during the MALDI process.

30 ESI mass spectrometry ("ESI-MS") has been of greater utility for studying non-covalent molecular interactions because, unlike the MALDI process, ESI-MS generates molecular ions with little to no fragmentation (Xavier *et al.*, 2000, *Trends Biotechnol.* 18(8):349-356). ESI-MS has been used to study the complexes formed by HIV Tat peptide and protein with the TAR RNA (Sannes-Lowery *et al.*, 1997, *Anal. Chem.* 69:5130-5135).

35 Fourier-transform ion cyclotron resonance ("FT-ICR") mass spectrometry provides high-resolution spectra, isotope-resolved precursor ion selection, and accurate mass

5 assignments (Xavier *et al.*, 2000, Trends Biotechnol. 18(8):349-356). FT-ICR has been used to study the interaction of aminoglycoside antibiotics with cognate and non-cognate RNAs (Hofstadler *et al.*, 1999, Anal. Chem. 71:3436-3440; Griffey *et al.*, 1999, Proc. Natl. Acad. Sci. USA 96:10129-10133). As true for all of the mass spectrometry methods discussed herein, FT-ICR does not require labeling of the target RNA or a test compound.

10 An advantage of mass spectroscopy is not only the elucidation of the structure of the test compound, but also the determination of the structure of the test compound bound to the preselected target RNA. Such information can enable the discovery of a consensus structure of a test compound that specifically binds to a preselected target RNA.

15 In a preferred embodiment, the structure of the test compound is determined by time of flight mass spectroscopy ("TOF-MS"). In time of flight methods of mass spectrometry, charged (ionized) molecules are produced in a vacuum and accelerated by an electric field into a time of flight tube or drift tube. The velocity to which the molecules may be accelerated is proportional to the accelerating potential, proportional to the charge of the molecule, and inversely proportional to the square of the mass of the molecule. The charged molecules travel, *i.e.*, "drift" down the TOF tube to a detector. The time taken for the molecules to travel down the tube may be interpreted as a measure of their molecular weight. Time-of-flight mass spectrometers have been developed for all of the major ionization 20 techniques such as, but limited to, electron impact ("EI"), infrared laser desorption ("IRLD"), plasma desorption ("PD"), fast atom bombardment ("FAB"), secondary ion mass spectrometry ("SIMS"), matrix-assisted laser desorption/ionization ("MALDI"), and electrospray ionization ("ESI").

25 4.6.2. NMR Spectroscopy

30 NMR spectroscopy can be used for elucidating the structure of the test compound on the isolated bead. NMR spectroscopy is a technique for identifying binding sites in target nucleic acids by qualitatively determining changes in chemical shift, specifically from distances measured using relaxation effects. Examples of NMR that can be used for the invention include, but are not limited to, one-dimensional NMR, two-dimensional NMR, correlation spectroscopy ("COSY"), and nuclear Overhauser effect ("NOE") spectroscopy. Such methods of structure determination of test compounds are well known to one of skill in the art.

35 Similar to mass spectroscopy, an advantage of NMR is the not only the elucidation of the structure of the test compound, but also the determination of the structure of the test compound bound to the preselected target RNA. Such information can enable the

discovery of a consensus structure of a test compound that specifically binds to a preselected target RNA.

4.6.3. Edman Degradation

In an embodiment wherein the library is a peptide library or a derivative thereof, Edman degradation can be used to determine the structure of the test compound. In one embodiment, a modified Edman degradation process is used to obtain compositional tags for proteins, which is described in U.S. Patent No. 6,277,644 to Farnsworth *et al.*, which is hereby incorporated by reference in its entirety. The Edman degradation chemistry is separated from amino acid analysis, circumventing the serial requirement of the conventional Edman process. Multiple cycles of coupling and cleavage are performed prior to extraction and compositional analysis of amino acids. The amino acid composition information is then used to search a database of known protein or DNA sequences to identify the sample protein. An apparatus for performing this method comprises a sample holder for holding the sample, a coupling agent supplier for supplying at least one coupling agent, a cleavage agent supplier for supplying a cleavage agent, a controller for directing the sequential supply of the coupling agents, cleavage agents, and other reagents necessary for performing the modified Edman degradation reactions, and an analyzer for analyzing amino acids.

20 In another embodiment, the method can be automated as described in U.S. Patent No. 5,565,171 to Dovichi *et al.*, which is hereby incorporated by reference in its entirety. The apparatus includes a continuous capillary connected between two valves that control fluid flow in the capillary. One part of the capillary forms a reaction chamber where the sample may be immobilized for subsequent reaction with reagents supplied through the valves. Another part of the capillary passes through or terminates in the detector portion of an analyzer such as an electrophoresis apparatus, liquid chromatographic apparatus or mass spectrometer. The apparatus may form a peptide or protein sequencer for carrying out the Edman degradation reaction and analyzing the reaction product produced by the reaction. The protein or peptide sequencer includes a reaction chamber for carrying out coupling and cleavage on a peptide or protein to produce derivatized amino acid residue, a conversion chamber for carrying out conversion and producing a converted amino acid residue and an analyzer for identifying the converted amino acid residue. The reaction chamber may be contained within one arm of a capillary and the conversion chamber is located in another arm of the capillary. An electrophoresis length of capillary is directly capillary coupled to the conversion chamber to allow electrophoresis separation of the converted amino acid residue

as it leaves the conversion chamber. Identification of the converted amino acid residue takes place at one end of the electrophoresis length of the capillary.

4.6.4. Vibrational Spectroscopy

Vibrational spectroscopy (e.g. infrared (IR) spectroscopy or Raman spectroscopy) can be used for elucidating the structure of the test compound on the isolated bead.

Infrared spectroscopy measures the frequencies of infrared light (wavelengths from 100 to 10,000 nm) absorbed by the test compound as a result of excitation of 10 vibrational modes according to quantum mechanical selection rules which require that absorption of light cause a change in the electric dipole moment of the molecule. The infrared spectrum of any molecule is a unique pattern of absorption wavelengths of varying intensity that can be considered as a molecular fingerprint to identify any compound.

15 Infrared spectra can be measured in a scanning mode by measuring the absorption of individual frequencies of light, produced by a grating which separates frequencies from a mixed-frequency infrared light source, by the test compound relative to a standard intensity (double-beam instrument) or pre-measured ('blank') intensity (single-beam instrument). In a preferred embodiment, infrared spectra are measured in a pulsed mode
20 (FT-IR) where a mixed beam, produced by an interferometer, of all infrared light frequencies is passed through or reflected off the test compound. The resulting interferogram, which may or may not be added with the resulting interferograms from subsequent pulses to increase the signal strength while averaging random noise in the electronic signal, is mathematically transformed into a spectrum using Fourier Transform or Fast Fourier Transform algorithms.

25 Raman spectroscopy measures the difference in frequency due to absorption of infrared frequencies of scattered visible or ultraviolet light relative to the incident beam. The incident monochromatic light beam, usually a single laser frequency, is not truly absorbed by the test compound but interacts with the electric field transiently. Most of the light scattered off the sample will be unchanged (Rayleigh scattering) but a portion of the scatter light will have frequencies that are the sum or difference of the incident and molecular vibrational frequencies. The selection rules for Raman (inelastic) scattering require a change in polarizability of the molecule. While some vibrational transitions are observable in both infrared and Raman spectrometry, most are observable only with one or the other technique. The Raman spectrum of any molecule is a unique pattern of absorption wavelengths of varying intensity that can be considered as a molecular fingerprint to identify any compound.

Raman spectra are measured by submitting monochromatic light to the sample, either passed through or preferably reflected off, filtering the Rayleigh scattered light, and detecting the frequency of the Raman scattered light. An improved Raman spectrometer is described in US Patent No. 5,786,893 to Fink *et al.*, which is hereby incorporated by reference.

Vibrational microscopy can be measured in a spatially resolved fashion to address single beads by integration of a visible microscope and spectrometer. A microscopic infrared spectrometer is described in U.S. Patent No. 5,581,085 to Reffner *et al.*, which is hereby incorporated by reference in its entirety. An instrument that simultaneously performs a microscopic infrared and microscopic Raman analysis on a sample is described in U.S. Patent No. 5,841,139 to Sostek *et al.*, which is hereby incorporated by reference in its entirety.

In one embodiment of the method, test compounds are synthesized on polystyrene beads doped with chemically modified styrene monomers such that each resulting bead has a characteristic pattern of absorption lines in the vibrational (IR or Raman) spectrum, by methods including but not limited to those described by Fenniri *et al.*, 2000, J. Am. Chem. Soc. 123:8151-8152. Using methods of split-pool synthesis familiar to one of skill in the art, the library of compounds is prepared so that the spectroscopic pattern of the bead identifies one of the components of the test compound on the bead. Beads that have been separated according to their ability to bind target RNA can be identified by their vibrational spectrum. In one embodiment of the method, appropriate sorting and binning of the beads during synthesis then allows identification of one or more further components of the test compound on any one bead. In another embodiment of the method, partial identification of the compound on a bead is possible through use of the spectroscopic pattern of the bead with or without the aid of further sorting during synthesis, followed by partial resynthesis of the possible compounds aided by doped beads and appropriate sorting during synthesis.

In another embodiment, the IR or Raman spectra of test compounds are examined while the compound is still on a bead, preferably, or after cleavage from bead, using methods including but not limited to photochemical, acid, or heat treatment. The test compound can be identified by comparison of the IR or Raman spectral pattern to spectra previously acquired for each test compound in the combinatorial library.

4.7. Secondary Biological Screens

5 The test compounds identified in the binding assay (for convenience referred to herein as a “lead” compound) can be tested for biological activity using host cells containing or engineered to contain the target RNA element coupled to a functional readout system. For example, the lead compound can be tested in a host cell engineered to contain the target RNA element controlling the expression of a reporter gene. In this example, the lead compounds are assayed in the presence or absence of the target RNA. Alternatively, a phenotypic or physiological readout can be used to assess activity of the target RNA in the presence and absence of the lead compound.

10

15 In one embodiment, the lead compound can be tested in a host cell engineered to contain the target RNA element controlling the expression of a reporter gene, such as, but not limited to, β -galactosidase, green fluorescent protein, red fluorescent protein, luciferase, chloramphenicol acetyltransferase, alkaline phosphatase, and β -lactamase. In a preferred embodiment, a cDNA encoding the target element is fused upstream to a reporter gene wherein translation of the reporter gene is repressed upon binding of the lead compound to the target RNA. In other words, the steric hindrance caused by the binding of the lead compound to the target RNA repressed the translation of the reporter gene. This method, termed the translational repression assay procedure (“TRAP”) has been demonstrated in *E. coli* and *S. cerevisiae* (Jain & Belasco, 1996, *Cell* 87(1):115-25; Huang & Schreiber, 1997, *Proc. Natl. Acad. Sci. USA* 94:13396-13401).

20

25 In another embodiment, a phenotypic or physiological readout can be used to assess activity of the target RNA in the presence and absence of the lead compound. For example, the target RNA may be overexpressed in a cell in which the target RNA is endogenously expressed. Where the target RNA controls expression of a gene product involved in cell growth or viability, the *in vivo* effect of the lead compound can be assayed by measuring the cell growth or viability of the target cell. Alternatively, a reporter gene can also be fused downstream of the target RNA sequence and the effect of the lead compound on reporter gene expression can be assayed.

30 Alternatively, the lead compounds identified in the binding assay can be tested for biological activity using animal models for a disease, condition, or syndrome of interest. These include animals engineered to contain the target RNA element coupled to a functional readout system, such as a transgenic mouse. Animal model systems can also be used to demonstrate safety and efficacy.

35 Compounds displaying the desired biological activity can be considered to be lead compounds, and will be used in the design of congeners or analogs possessing useful

pharmacological activity and physiological profiles. Following the identification of a lead compound, molecular modeling techniques can be employed, which have proven to be useful in conjunction with synthetic efforts, to design variants of the lead that can be more effective.

5 These applications may include, but are not limited to, Pharmacophore Modeling (*cf.* Lamothe, *et al.* 1997, *J. Med. Chem.* 40: 3542; Mottola *et al.* 1996, *J. Med. Chem.* 39: 285; Beusen *et al.* 1995, *Biopolymers* 36: 181; P. Fossa *et al.* 1998, *Comput. Aided Mol. Des.* 12: 361), QSAR development (*cf.* Siddiqui *et al.* 1999, *J. Med. Chem.* 42: 4122; Barreca *et al.* 1999 *Bioorg. Med. Chem.* 7: 2283; Kroemer *et al.* 1995, *J. Med. Chem.* 38: 4917; Schaal *et al.* 10 2001, *J. Med. Chem.* 44: 155; Buolamwini & Assefa 2002, *J. Mol. Chem.* 45: 84), Virtual docking and screening/scoring (*cf.* Anzini *et al.* 2001, *J. Med. Chem.* 44: 1134; Faaland *et al.* 2000, *Biochem. Cell. Biol.* 78: 415; Silvestri *et al.* 2000, *Bioorg. Med. Chem.* 8: 2305; J. Lee *et al.* 2001, *Bioorg. Med. Chem.* 9: 19), and Structure Prediction using RNA structural 15 programs including, but not limited to mFold (as described by Zuker *et al.* Algorithms and Thermodynamics for RNA Secondary Structure Prediction: A Practical Guide in RNA Biochemistry and Biotechnology pp. 11-43, J. Barciszewski & B.F.C. Clark, eds. (NATO ASI Series, Kluwer Academic Publishers, 1999) and Mathews *et al.* 1999 *J. Mol. Biol.* 288: 911-940); RNAmotif (Macke *et al.* 2001, *Nucleic Acids Res.* 29: 4724-4735; and the Vienna RNA package (Hofacker *et al.* 1994, *Monatsh. Chem.* 125: 167-188)).

20 Further examples of the application of such techniques can be found in several review articles, such as Rotivinen *et al.*, 1988, *Acta Pharmaceutical Fennica* 97:159-166; Ripka, 1998, *New Scientist* 54-57; McKinlay & Rossmann, 1989, *Annu. Rev. Pharmacol. Toxicol.* 29:111-122; Perry & Davies, QSAR: Quantitative Structure-Activity Relationships in Drug Design pp. 189-193 (Alan R. Liss, Inc. 1989); Lewis & Dean, 1989, *Proc. R. Soc. Lond.* 236:125-140 and 141-162; Askew *et al.*, 1989, *J. Am. Chem. Soc.* 111:1082-1090.

25 Molecular modeling tools employed may include those from Tripos, Inc., St. Louis, Missouri (e.g., Sybyl/UNITY, CONCORD, DiverseSolutions), Accelrys, San Diego, California (e.g., Catalyst, Wisconsin Package {BLAST, etc.}), Schrodinger, Portland, Oregon (e.g., QikProp, QikFit, Jaguar) or other such vendors as BioDesign, Inc. (Pasadena, California), Allelix, Inc. (Mississauga, Ontario, Canada), and Hypercube, Inc. (Cambridge, Ontario, Canada), and may 30 include privately designed and/or "academic" software (e.g. RNAMotif, mFOLD). These application suites and programs include tools for the atomistic construction and analysis of structural models for drug-like molecules, proteins, and DNA or RNA and their potential interactions. They also provide for the calculation of important physical properties, such as solubility estimates, permeability metrics, and empirical measures of molecular 35 "druggability" (e.g., Lipinski "Rule of 5" as described by Lipinski *et al.* 1997, *Adv. Drug*

Delivery Rev. 23: 3-25). Most importantly, they provide appropriate metrics and statistical modeling power (such as the patented CoMFA technology in Sybyl as described in US Patents 6,240,374 and 6,185,506) to develop Quantitative Structural Activity Relationships (QSARs) which are used to guide the synthesis of more efficacious clinical development candidates while improving desirable physical properties, as determined by results from the aforementioned secondary screening protocols.

4.8. Use of Identified Compounds That Bind RNA to Treat/Prevent Disease

Biologically active compounds identified using the methods of the invention or a pharmaceutically acceptable salt thereof can be administered to a patient, preferably a mammal, more preferably a human, suffering from a disease whose progression is associated with a target RNA:host cell factor interaction *in vivo*. In certain embodiments, such compounds or a pharmaceutically acceptable salt thereof is administered to a patient, preferably a mammal, more preferably a human, as a preventative measure against a disease associated with an RNA:host cell factor interaction *in vivo*.

In one embodiment, "treatment" or "treating" refers to an amelioration of a disease, or at least one discernible symptom thereof. In another embodiment, "treatment" or "treating" refers to an amelioration of at least one measurable physical parameter, not necessarily discernible by the patient. In yet another embodiment, "treatment" or "treating" refers to inhibiting the progression of a disease, either physically, *e.g.*, stabilization of a discernible symptom, physiologically, *e.g.*, stabilization of a physical parameter, or both. In yet another embodiment, "treatment" or "treating" refers to delaying the onset of a disease.

In certain embodiments, the compound or a pharmaceutically acceptable salt thereof is administered to a patient, preferably a mammal, more preferably a human, as a preventative measure against a disease associated with an RNA:host cell factor interaction *in vivo*. As used herein, "prevention" or "preventing" refers to a reduction of the risk of acquiring a disease. In one embodiment, the compound or a pharmaceutically acceptable salt thereof is administered as a preventative measure to a patient. According to this embodiment, the patient can have a genetic predisposition to a disease, such as a family history of the disease, or a non-genetic predisposition to the disease. Accordingly, the compound and pharmaceutically acceptable salts thereof can be used for the treatment of one manifestation of a disease and prevention of another.

When administered to a patient, the compound or a pharmaceutically acceptable salt thereof is preferably administered as component of a composition that optionally comprises a pharmaceutically acceptable vehicle. The composition can be

5 administered orally, or by any other convenient route, for example, by infusion or bolus injection, by absorption through epithelial or mucocutaneous linings (e.g., oral mucosa, rectal, and intestinal mucosa, *etc.*) and may be administered together with another biologically active agent. Administration can be systemic or local. Various delivery systems are known, *e.g.*, encapsulation in liposomes, microparticles, microcapsules, capsules, *etc.*, and can be used to administer the compound and pharmaceutically acceptable salts thereof.

10 Methods of administration include but are not limited to intradermal, intramuscular, intraperitoneal, intravenous, subcutaneous, intranasal, epidural, oral, sublingual, intranasal, intracerebral, intravaginal, transdermal, rectally, by inhalation, or 15 topically, particularly to the ears, nose, eyes, or skin. The mode of administration is left to the discretion of the practitioner. In most instances, administration will result in the release of the compound or a pharmaceutically acceptable salt thereof into the bloodstream.

15 In specific embodiments, it may be desirable to administer the compound or a pharmaceutically acceptable salt thereof locally. This may be achieved, for example, and not by way of limitation, by local infusion during surgery, topical application, *e.g.*, in conjunction with a wound dressing after surgery, by injection, by means of a catheter, by means of a suppository, or by means of an implant, said implant being of a porous, non-porous, or gelatinous material, including membranes, such as sialastic membranes, or fibers.

20 In certain embodiments, it may be desirable to introduce the compound or a pharmaceutically acceptable salt thereof into the central nervous system by any suitable route, including intraventricular, intrathecal and epidural injection. Intraventricular injection may be facilitated by an intraventricular catheter, for example, attached to a reservoir, such as an Ommaya reservoir.

25 Pulmonary administration can also be employed, *e.g.*, by use of an inhaler or nebulizer, and formulation with an aerosolizing agent, or via perfusion in a fluorocarbon or synthetic pulmonary surfactant. In certain embodiments, the compound and pharmaceutically acceptable salts thereof can be formulated as a suppository, with traditional binders and vehicles such as triglycerides.

30 In another embodiment, the compound and pharmaceutically acceptable salts thereof can be delivered in a vesicle, in particular a liposome (see Langer, 1990, *Science* 249:1527-1533; Treat *et al.*, in *Liposomes in the Therapy of Infectious Disease and Cancer*, Lopez-Berestein and Fidler (eds.), Liss, New York, pp. 353-365 (1989); Lopez-Berestein, *ibid.*, pp. 317-327; see generally *ibid.*).

35 In yet another embodiment, the compound and pharmaceutically acceptable salts thereof can be delivered in a controlled release system (see, *e.g.*, Goodson, in *Medical*

Applications of Controlled Release, *supra*, vol. 2, pp. 115-138 (1984)). Other controlled-release systems discussed in the review by Langer, 1990, *Science* 249:1527-1533) may be used. In one embodiment, a pump may be used (see Langer, *supra*; Sefton, 1987, *CRC Crit. Ref. Biomed. Eng.* 14:201; Buchwald *et al.*, 1980, *Surgery* 88:507 Saudek *et al.*, 1989, *N. Engl. J. Med.* 321:574). In another embodiment, polymeric materials can be used (see Medical Applications of Controlled Release, Langer and Wise (eds.), *CRC Pres.*, Boca Raton, Florida (1974); Controlled Drug Bioavailability, Drug Product Design and Performance, Smolen and Ball (eds.), *Wiley*, New York (1984); Ranger and Peppas, 1983, *J. Macromol. Sci. Rev. Macromol. Chem.* 23:61; see also Levy *et al.*, 1985, *Science* 228:190; During *et al.*, 1989, *Ann. Neurol.* 25:351; Howard *et al.*, 1989, *J. Neurosurg.* 71:105). In yet another embodiment, a controlled-release system can be placed in proximity of a target RNA of the compound or a pharmaceutically acceptable salt thereof, thus requiring only a fraction of the systemic dose.

Compositions comprising the compound or a pharmaceutically acceptable salt thereof ("compound compositions") can additionally comprise a suitable amount of a pharmaceutically acceptable vehicle so as to provide the form for proper administration to the patient.

In a specific embodiment, the term "pharmaceutically acceptable" means approved by a regulatory agency of the Federal or a state government or listed in the U.S. Pharmacopeia or other generally recognized pharmacopeia for use in animals, mammals, and more particularly in humans. The term "vehicle" refers to a diluent, adjuvant, excipient, or carrier with which a compound of the invention is administered. Such pharmaceutical vehicles can be liquids, such as water and oils, including those of petroleum, animal, vegetable or synthetic origin, such as peanut oil, soybean oil, mineral oil, sesame oil and the like. The pharmaceutical vehicles can be saline, gum acacia, gelatin, starch paste, talc, keratin, colloidal silica, urea, and the like. In addition, auxiliary, stabilizing, thickening, lubricating and coloring agents may be used. When administered to a patient, the pharmaceutically acceptable vehicles are preferably sterile. Water is a preferred vehicle when the compound of the invention is administered intravenously. Saline solutions and aqueous dextrose and glycerol solutions can also be employed as liquid vehicles, particularly for injectable solutions. Suitable pharmaceutical vehicles also include excipients such as starch, glucose, lactose, sucrose, gelatin, malt, rice, flour, chalk, silica gel, sodium stearate, glycerol monostearate, talc, sodium chloride, dried skim milk, glycerol, propylene, glycol, water, ethanol and the like. Compound compositions, if desired, can also contain minor amounts of wetting or emulsifying agents, or pH buffering agents.

Compound compositions can take the form of solutions, suspensions, emulsion, tablets, pills, pellets, capsules, capsules containing liquids, powders, sustained-release formulations, suppositories, emulsions, aerosols, sprays, suspensions, or any other form suitable for use. In one embodiment, the pharmaceutically acceptable vehicle is a capsule (see e.g., U.S. Patent No. 5,698,155). Other examples of suitable pharmaceutical vehicles are described in Remington's Pharmaceutical Sciences, Alfonso R. Gennaro, ed., Mack Publishing Co. Easton, PA, 19th ed., 1995, pp. 1447 to 1676, incorporated herein by reference.

In a preferred embodiment, the compound or a pharmaceutically acceptable salt thereof is formulated in accordance with routine procedures as a pharmaceutical composition adapted for oral administration to human beings. Compositions for oral delivery may be in the form of tablets, lozenges, aqueous or oily suspensions, granules, powders, emulsions, capsules, syrups, or elixirs, for example. Orally administered compositions may contain one or more agents, for example, sweetening agents such as fructose, aspartame or saccharin; flavoring agents such as peppermint, oil of wintergreen, or cherry; coloring agents; and preserving agents, to provide a pharmaceutically palatable preparation. Moreover, where in tablet or pill form, the compositions can be coated to delay disintegration and absorption in the gastrointestinal tract thereby providing a sustained action over an extended period of time. Selectively permeable membranes surrounding an osmotically active driving compound are also suitable for orally administered compositions. In these later platforms, fluid from the environment surrounding the capsule is imbibed by the driving compound, which swells to displace the agent or agent composition through an aperture. These delivery platforms can provide an essentially zero order delivery profile as opposed to the spiked profiles of immediate release formulations. A time delay material such as glycerol monostearate or glycerol stearate may also be used. Oral compositions can include standard vehicles such as mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate, and the like. Such vehicles are preferably of pharmaceutical grade. Typically, compositions for intravenous administration comprise sterile isotonic aqueous buffer. Where necessary, the compositions may also include a solubilizing agent.

In another embodiment, the compound or a pharmaceutically acceptable salt thereof can be formulated for intravenous administration. Compositions for intravenous administration may optionally include a local anesthetic such as lignocaine to lessen pain at the site of the injection. Generally, the ingredients are supplied either separately or mixed together in unit dosage form, for example, as a dry lyophilized powder or water-free

concentrate in a hermetically sealed container such as an ampoule or sachette indicating the quantity of active agent. Where the compound or a pharmaceutically acceptable salt thereof is to be administered by infusion, it can be dispensed, for example, with an infusion bottle 5 containing sterile pharmaceutical grade water or saline. Where the compound or a pharmaceutically acceptable salt thereof is administered by injection, an ampoule of sterile water for injection or saline can be provided so that the ingredients may be mixed prior to administration.

The amount of a compound or a pharmaceutically acceptable salt thereof that 10 will be effective in the treatment of a particular disease will depend on the nature of the disease, and can be determined by standard clinical techniques. In addition, *in vitro* or *in vivo* assays may optionally be employed to help identify optimal dosage ranges. The precise dose to be employed will also depend on the route of administration, and the seriousness of the disease, and should be decided according to the judgment of the practitioner and each 15 patient's circumstances. However, suitable dosage ranges for oral administration are generally about 0.001 milligram to about 200 milligrams of a compound or a pharmaceutically acceptable salt thereof per kilogram body weight per day. In specific preferred embodiments of the invention, the oral dose is about 0.01 milligram to about 100 milligrams per kilogram body weight per day, more preferably about 0.1 milligram to about 20 75 milligrams per kilogram body weight per day, more preferably about 0.5 milligram to 5 milligrams per kilogram body weight per day. The dosage amounts described herein refer to total amounts administered; that is, if more than one compound is administered, or if a compound is administered with a therapeutic agent, then the preferred dosages correspond to the total amount administered. Oral compositions preferably contain about 10% to about 25 95% active ingredient by weight.

Suitable dosage ranges for intravenous (i.v.) administration are about 0.01 milligram to about 100 milligrams per kilogram body weight per day, about 0.1 milligram to about 35 milligrams per kilogram body weight per day, and about 1 milligram to about 10 milligrams per kilogram body weight per day. Suitable dosage ranges for intranasal 30 administration are generally about 0.01 pg/kg body weight per day to about 1 mg/kg body weight per day. Suppositories generally contain about 0.01 milligram to about 50 milligrams of a compound of the invention per kilogram body weight per day and comprise active ingredient in the range of about 0.5% to about 10% by weight.

Recommended dosages for intradermal, intramuscular, intraperitoneal, 35 subcutaneous, epidural, sublingual, intracerebral, intravaginal, transdermal administration or administration by inhalation are in the range of about 0.001 milligram to about 200

milligrams per kilogram of body weight per day. Suitable doses for topical administration are in the range of about 0.001 milligram to about 1 milligram, depending on the area of administration. Effective doses may be extrapolated from dose-response curves derived from 5 *in vitro* or animal model test systems. Such animal models and systems are well known in the art.

The compound and pharmaceutically acceptable salts thereof are preferably assayed *in vitro* and *in vivo*, for the desired therapeutic or prophylactic activity, prior to use in humans. For example, *in vitro* assays can be used to determine whether it is preferable to 10 administer the compound, a pharmaceutically acceptable salt thereof, and/or another therapeutic agent. Animal model systems can be used to demonstrate safety and efficacy.

A variety of compounds can be used for treating or preventing diseases in mammals. Types of compounds include, but are not limited to, peptides, peptide analogs including peptides comprising non-natural amino acids, e.g., D-amino acids, phosphorous 15 analogs of amino acids, such as α -amino phosphonic acids and α -amino phosphinic acids, or amino acids having non-peptide linkages, nucleic acids, nucleic acid analogs such as phosphorothioates or peptide nucleic acids ("PNAs"), hormones, antigens, synthetic or naturally occurring drugs, opiates, dopamine, serotonin, catecholamines, thrombin, 20 acetylcholine, prostaglandins, organic molecules, pheromones, adenosine, sucrose, glucose, lactose and galactose.

5. EXAMPLE: THERAPEUTIC TARGETS

The therapeutic targets presented herein are by way of example, and the present invention is not to be limited by the targets described herein. The therapeutic targets 25 presented herein as DNA sequences are understood by one of skill in the art that the sequences can be converted to RNA sequences.

5.1. Tumor Necrosis Factor Alpha ("TNF- α ")

GenBank Accession # X01394:

30 1 gcagaggacc agctaagagg gagagaagca actacagacc cccctgaaa acaaccctca
61 gacgccacat cccctgacaa gctgccaggc aggttcttt ccttcacat actgaccac
121 ggctccaccc tctctccct ggaaaggaca ccatgagcac tgaaaggat atccgggacg
181 tggagctggc cgaggaggcg ctcccaaga agacaggggg gccccaggc tccaggcggt
241 gcttgttct cagccttc tcccttctga tcgtggcagg cggcaccacg ctcttctgcc
301 tgctgcactt tggagtgtac ggccccaga gggaaaggtt ccccaaggac ctcttcataa
361 tcagccctct ggcccaggca gtcagatcat ctctcgaac cccgagtgc aagcctgttag

421 cccatgttgt agcaaaccct caagctgagg ggcagctcca gtggctgaac cgccgggcca
 481 atgcccctcct ggccaatggc gtggagctga gagataacca gctgggtggc ccatcagagg
 541 gcctgtacct catctactcc caggccctct tcaaggggca aggctgcccc tccacccatg
 601 tgcctcctac ccacaccatc agccgcacg cgcgtcctca ccagaccaag gtcaacccct
 661 tctctgcccataa caagagcccc tgccagaggg agacccaga gggggctgag gccaaaggcc
 721 ggtatgagcc catctatctg ggaggggtct tccagctgga gaagggtgac cgactcagcg
 781 ctgagatcaa tcggcccgac tatttcgact ttgcgcagtc tgggcaggc tacitggga
 841 tcattggccct gtgaggagga cgaacatcca accttccaa acgcctcccc tgcccaatc
 901 ctttattac cccctccctc agacaccctc aaccccttct ggctcaaaaa gagaattggg
 961 ggcttagggt cggaacccaa gcttagaact ttaagcaaca agaccaccac ttgcaccc
 1021 gggattcagg aatgtgtggc ctgcacatgt aattgtggc aaccactaag aattcaaact
 1081 gggcctcca gaactcactg gggctacag cttgtatccc tgacatctgg aatctggaga
 1141 ccagggagcc ttgggtctg gccagaatgc tgcaggactt gagaagaccc cacctagaaa
 1201 ttgacacaag tggaccccttgc gccttcctct ctcctccatgt ttccagactt ctttgcac
 1261 cggagcccaag ccctcccat ggagccagct ccctctatgt atgttgac ttgtgattat
 1321 ttattattta ttattattttt atttattttc agatgaatgt atttattttgg gagaccgggg
 1381 tatcctgggg gaccaatgt aggagctgcc ttggctcaga catgtttcc gtggaaaacgg
 1441 agctgaacaa taggctgttc ccatgttagcc ccctggcctc tgccttc tttgattatg
 1501 tttttaaaaa tatttatctg attaagttgt ctaaacaatgt ctgattttgtt gaccaactgt
 1561 cactcatgtc tgaggctctg ctcccccaggaggtgtgtc tgtaatcgcc ctactattca
 1621 gtggcgagaa ataaagtgg ctt (SEQ ID NO: 6)

General Target Regions:

- 25 (1) 5' Untranslated Region - nts 1 - 152
 (2) 3' Untranslated Region - nts 852 - 1643

Initial Specific Target Motif:

Group I AU-Rich Element (ARE) Cluster in 3' untranslated region
 30 5' AUUUAUUUAUUUAUUUAUUA 3' (SEQ ID NO: 1)

5.2. Granulocyte-macrophage Colony Stimulating Factor (“GM-CSF”)

GenBank Accession # NM_000758:

1 gctggaggat gtggctgcag agcctgctgc tcttggcac tgcgtgc acgtatctgt
 61 caccggcccg ctgcggccagc cccagcacgc agccctggga gcatgtgaat gccatccagg
 35 121 aggcccgccg tctccctgaac ctgagtagag acactgctgc tgagatgaat gaaacagtag

181 aagtcatctc agaaatgtt gacctccagg agccgaccctg cctacagacc cgccctggagc
 241 tgtacaagca gggcctgcgg ggcagccca ccaagctcaa gggcccttg accatgatgg
 301 ccagccacta caagcagcac tgccctccaa ccccgaaac ttccctgtgca acccagacta
 5 361 tcaccttga aagttcaaa gagaacctga aggacttct gcttgcata cccttgact
 421 gctgggagcc agtccaggag tgagaccggc cagatgaggc tggccaagcc ggggagctgc
 481 tctctcatga aacaagagct agaaactcag gatggtcata ttggagggac caaggggtgg
 541 gccacagcca tggtggagt ggcctggacc tgccctggc cacactgacc ctgatacagg
 601 catggcagaa gaatggaaat attttatact gacagaaatc agtaatattt atatatttat
 10 661 attttaaaaa tattttattt ttattttatt taagttcata ttccatattt attcaagatg
 721 tttaccgta ataatttata ttaaaaatat gcttct (SEQ ID NO: 7)

GenBank Accession # XM_003751:

1 tctggaggat gtggctgcag agcctgcgtc tcttggcac tggccctgc agcatctcg
 15 61 cacccgcggc ctcgcggcgc cccagcacgc agccctggga gcatgtaat gccatccagg
 121 aggcccgccg tctcctgaac ctgagtagag acactgctgc tgagatgaat gaaacagtag
 181 aagtcatctc agaaatgtt gacctccagg agccgaccctg cctacagacc cgccctggagc
 241 tgtacaagca gggcctgcgg ggcagccca ccaagctcaa gggcccttg accatgatgg
 301 ccagccacta caagcagcac tgccctccaa ccccgaaac ttccctgtgca acccagacta
 20 361 tcaccttga aagttcaaa gagaacctga aggacttct gcttgcata cccttgact
 421 gctgggagcc agtccaggag tgagaccggc cagatgaggc tggccaagcc ggggagctgc
 481 tctctcatga aacaagagct agaaactcag gatggtcata ttggagggac caaggggtgg
 541 gccacagcca tggtggagt ggcctggacc tgccctggc cacactgacc ctgatacagg
 601 catggcagaa gaatggaaat attttatact gacagaaatc agtaatattt atatatttat
 25 661 attttaaaaa tattttattt ttattttatt taagttcata ttccatattt attcaagatg
 721 tttaccgta ataatttata ttaaaaatat gcttct (SEQ ID NO: 8)

General Target Regions:

- 30 (1) 5' Untranslated Region - nts 1 - 32
 (2) 3' Untranslated Region - nts 468 - 789

Initial Specific Target Motif:

Group I AU-Rich Element (ARE) Cluster in 3' untranslated region
 35 5' AUUUUUUUUUUUUUUUUUUA 3' (SEQ ID NO: 1)

5.3. Interleukin 2 (“IL-2”)

GenBank Accession # U25676:

1 atcactctctttaatcactactcacattaa cctcaactcc tgccacaatg tacaggatgc
5 61 aactccgtcttgcattgca ctaattcttg cacttgtcac aaacagtgcac cctacttcaa
121 gtgcacaaa gaaaacaaaag aaaacacagc tacaactgga gcatttacig ctgattttac
181 agatgatttttaatggaatt aataattaca agaatcccaa actcaccagg atgctcacat
241 ttaagtttaatgcccag aaggccacag aactgaaaca gcttcagtgt ctagaagaag
301 aactcaaacc tctggagggaa gtgctgaatt tagctcaaag caaaaacttt cacttaagac
361 ccagggactt aatcagcaat atcaacgtaa tagttctgga actaaaggga tctgaaacaa
421 cattcaigtgttgaatatgca gatgagacag caaccattgtt agaattctg aacagatgg
481 ttaccccttgc tcaaagcatc atctcaacac taacttgata attaagtgtctccacttaa
541 aacatatcag gccttctatt tatttttttaa aatattttaaa ttttattttt attgttgaat
601 gtatgggtgc tacctattgtt aactattattt cttaatctta aaactataaa tatggatctt
661 ttatgattcttttgtaagc cctaggggctt cttaaatggtttaccttattt tatccaaaaa
721 atatttttaatgttgtaa tgtaatataat agtatactatg tagattggttt agtaaaaacta
781 ttaataaaat ttgataaataataaaaaaaa aaacaaaaaaa aaaaa (SEQ ID NO: 9)

General Target Regions:

- 20 (1) 5' Untranslated Region - nts 1 - 47
(2) 3' Untranslated Region - nts 519- 825

Initial Specific Target Motifs:

25 Group III AU-Rich Element (ARE) Cluster in 3' untranslated region
5' NAUUUAUUUAUUUAN 3' (SEQ ID NO: 10)

5.4. Interleukin 6 (“IL-6”)

GenBank Accession # NM 000600:

1 ttctgcctc gagcccacccg ggaacgaaag agaagctcta ttcgcctcc aggagccag
30 61 ctatgaactc cttctccaca agcgccttcg gtccagttgc cttctccctg gggctgtcc
121 tggtgttgcg tgctgccttc cctgccccag tacccccagg agaagattcc aaagatgtag
181 ccgccccaca cagacagcca ctcacactt cagaacgaat tgacaaacaa attcggtaca
241 tcctcgacgg catctcagcc ctgagaaagg agacatgtaa caagagtaac atgtgtgaaa
301 gcagcaaaga ggcactggca gaaaacaacc tgaaccctcc aaagatggct gaaaaagatg
35 361 gatgcttcca atctggattc aatgaggaga cttgcctgggt gaaaatcatc actggcttt
421 tggagtttga ggtataccta ggtacccctcc agaacagatt tgagagttgtt gaggaacaag

481 ccagagctgt gcagatgagt acaaaagtcc tcatccagtt cctgcagaaa aaggcaaaga
 541 atcttagatgc aataaccacc cctgacccaa ccacaaatgc cagcctgctg acgaagctgc
 601 aggacacagaa ccagtggctg caggacatga caactcatct cattctgcgc agctttaagg
 5 661 agttcctgca gtccagcctg agggctttc ggcaaatgtt gcatggcac ctcagattgt
 721 ttttttaat gggcattccct tcttctggc agaaacctgt ccactggca cagaacttat
 781 gttttctct atggagaact aaaagtatga gcgttaggac actattttaa ttattttaa
 841 ttatataata tttaaatatg tgaagctgag ttaattttatg taagtcatat ttatatttt
 901 aagaagtacc acttgaaca ttttatgtat tagttttgaa ataataatgg aaagtggcta
 10 961 tgcagttga atatccttg ttccagagcc agatcatttgc ttggaaagtgc taggcttacc
 1021 tcaaataaat ggctaactta tacatatttt taaagaataa tttatattgt attttatataa
 1081 tgtataaatg gttttatac caataaatgg cattttaaaa aattc (SEQ ID NO: 11)

General Target Regions:

- 15 (1) 5' Untranslated Region - nts 1 - 62
 (2) 3' Untranslated Region - nts 699 - 1125

Initial Specific Target Motifs:

Group III AU-Rich Element (ARE) Cluster in 3' untranslated region
 20 5' NAUUUAUUAUUAUUAN 3' (SEQ ID NO: 10)

5.5. Vascular Endothelial Growth Factor ("VEGF")

GenBank Accession # AF022375:

1 aagagctcca gagagaagtc gaggaagaga gagacggggt cagagagagc ggcggggcgt
 25 61 gcgagcagcg aaagcgacag gggcaaagtgc agtgcacgtc ttttgggggt gaccggcgga
 121 gcgccggcgtg agccctcccc cttgggatcc cgcagctgac cagtcgcgc gacggacaga
 181 cagacagaca ccccccggc ccccgatcac cacctctcc cccggccggcg gccggacagtgc
 241 gacgcggcg cgagccgcgg gcagggcccg gagccgcggcc cccggaggcg ggtggagggg
 30 301 gtcggagctc gcggcgctgc actgaaactt ttctgtccaaac ttctggctg ttctcgcttc
 361 ggaggagccg tggccgcgc gggggaaagcc gagccgagcg gagccgcgag aagtgcgtac
 421 tcggccggg aggagccgca gccggaggag ggggaggagg aagaagagaa ggaagaggag
 481 agggggccgc agtggcgact cggcgctgg aagccgggt catggacggg tgaggccggc
 541 gtgtgcgcag acagtgcctcc agcgccgcgc ctccccagcc ctggccggc ctcggccgg
 601 gaggaagagt agctgcgcga ggccgcgagg agagccggcc gccccacagc ccgagccgga
 661 gagggacgcg agccgcgcgc cccggcggg cctccgaaac catgaacttt ctgcgtctt
 721 gggtgcattt gggcccttgcc ttgtgtctt acctccacca tgccaaatgg tcccaggctg

781 cacccatggc agaaggagga gggcagaatc atcacgaagt ggtgaagttc atggatgtct
841 atcagcgcag ctactgccat ccaatcgaga ccctgggaa catctccag gagtaccctg
901 atgagatcgta gtacatcttc aagccatcct gtgtccctt gatgcgtgc gggggctgt
961 ccaatgacga gggcctggag tgtgtgccca ctgaggagtc caacatcacc atgcagatta
1021 tgcggatcaa acctcacaa ggccagcaca taggagagat gagttccta cagcacaaca
1081 aatgtgaatg cagaccaaag aaagatagag caagacaaga aaatccctgt gggccttgct
1141 cagagcggag aaagcatttg ttgtacaag atccgcagac gtgtaaatgt tcctgcaaaa
1201 acacacactc gcgttgcagc gcgaggcgc ttgagttaa cgaacgtact tgcagatgtg
1261 acaagccgag gcggtgagcc gggcaggagg aaggagcctc ctcagggtt tcgggaacca
1321 gatctctc cagggaaagac tgatacagaa cgatcgatac agaaaccacg ctggccac
1381 cacaccatca ccatcgacag aacagtcctt aatccagaaa cctgaaatga aggaagagga
1441 gactctgcgc agagcactt gggtccggag ggcgagactc cggcggaaacg attccgggc
1501 gggtaccca gcacggtccc tcttggattt ggattcgcca ttttttttt cttgtgtcta
1561 aatcaccgag cccggaagat tagagagttt tattctggg attcctgttag acacacccac
1621 ccacatacat acatttatat atatatata tatataaaaaataaaatata tatctctatt
1681 ttatataaaaaataaaatata tattttttt taaaatataac agtgctaatg ttattgggt
1741 cttcaactggc tgtagttgac tgctgtggc ttgagttggg agggaaatgt tcccaactcag
1801 atccgtacag ggaagaggag gagatgagag actctggcat gatctttttt ttgtcccaact
1861 tggtggggcc agggcctct cccctgccc agaatgtgc agggcaggc atgggggcaaa
1921 atatgaccca gttttggaa caccgacaaa cccagccctg ggcgtgagcc tctctacccc
1981 aggtcagacg gacagaaaga caaatcacag gtccggat gaggacacccg gctctgacca
2041 ggagttggg gagttcagg acattgtgt gctttgggaa ttccctccac atgtgcacg
2101 cgcacatcgcc ccccaggggc actgcctggc agattcaggc gcctggggc cttcgctt
2161 ctctcacatg cttctgagtt gcccaggagg ccactggcag atgtccggc gaagagaaga
2221 gacacattgt tggaagaagc agcccatgac aegccctt cctggactc gcccctatcc
2281 tcttcctgt ccccttcctg gggtgcagcc taaaaggacc tatgtccctca caccattgaa
2341 accaactgtt ctgtcccccc agggaaacctg gtgtgtgtg tgtagtggt tgacccct
2401 ccatccctg gtcctccct tccctcccg aggcacagag agacaggcga ggttccacgt
2461 gcccattgtg gaggcagaga aaagagaaaag tgttttatatacggtactt ttaataatcc
2521 ctttttaattt agaaattttaga acagtttaattt taattaaaga gtagggttttt ttttcgtat
2581 tcttggtaa tatttaattt caactattt tgagatgtat cttttgtct ctcttgctct
2641 ctattttgtt ccggttttt tgataaaaaat tcatgttcc aatctctc tccctgtatcg
2701 gtgacagtca ctatgttac ttgaacagat atttaattttt gctaacactc agctctgccc
2761 tccccatgcc cctggctccc cagcacacat tcccttgaaa gagggtttca atatacatct
2821 acatactata tataatttgg gcaactgtt tttgtgtta tataatata tataatgttta

2881 tgtatataatg tgatccgtaa aaaataaaaca tcgctattct gtttttata tggtaaaacc
 2941 aaacaagaaa aaatagagaa ttctacatac taaatctctc tcctttta attttaatat
 3001 ttgttatcat ttatttattg gtgctactgt ttatccgtaa taattgtggg gaaaagat
 5 3061 taacatcagc tcttgcgtc tagtgcagtt ttgcgagata ttccgtagta catatttatt
 3121 tttaaacaac gacaaagaaa tacagatata tcttaaaaaa aaaaaa (SEQ ID NO: 12)

General Target Regions:

- 10 (1) 5' Untranslated Region - nts 1 - 701
 (2) 3' Untranslated Region - nts 1275 - 3166

Initial Specific Target Motifs:

- 15 (1) Internal Ribosome Entry Site (IRES) in 5' untranslated region nts 513 -704
 5'CCGGGCUCAUGGACGGGUGAGGCAGCGGGUGUGCGCAGACAGUG
 CUCCAGCGCGCGCUCCCCAGCCCUGGCCGGCCUCGGGCCGG
 AGGAAGAGUAGCUCGCCGAGGCCGAGGGAGAGCAGGGCCGCCCC
 ACAGCCCGAGCCGGAGAGGGACGCGAGCCGCGCCCCGGUCGG
 GCCUCCGAAACCAUGAACUUUCUGCUGCUUUGGGUGCAUUGGAG
 CCUUGCCUUGCUGCUCUACCUCCACCAUG 3' (SEQ ID NO: 13)
 20 (2) Group III AU-Rich Element (ARE) Cluster in 3' untranslated region
 5' NAUUUAUUUAUUUAN 3' (SEQ ID NO: 10)

5.6. Human Immunodeficiency Virus I ("HIV-1")

GenBank Accession # NC_001802:

25 1 ggctctctg gtttagaccag atctgaggct gggaggcttc tggcttaacta gggaaacccac
 61 tgcttaagcc tcaataaaacg ttgcctttag tgcttcaagt agtgtgtgcc cgtctgtgt
 121 gtgactctgg taactagaga tccctcagac ccttttagtc agtgtggaaa atctctagca
 181 gtggcgcccg aacagggacc tgaagcgaa agggaaacca gaggagctct ctgcacgcag
 241 gactcggctt gctgaaggcgc gcacggcaag aggcgagggg cggcgactgg tgagtacgcc
 301 aaaaattttt actagcggag gctagaagga gagagatggg tgcgagagcg tcagtattaa
 361 gcgggggaga attagatcga tggaaaaaaa ttccggtaag gccaggggaa aagaaaaat
 421 ataaattaaa acatatagtt tggcaagca gggagctaga acgattcgca gttaatcctg
 481 gcctgtttaga aacatcagaa ggctgttagac aaatactggg acagctacaa ccattcccttc
 541 agacaggatc agaagaactt agatcattat ataatacagt agcaaccctc tattgtgtgc
 601 atcaaaggat agagataaaa gacaccaagg aagctttaga caagatagag gaagagcaaa
 35 661 acaaaagtaa gaaaaaagca cagcaaggcag cagctgacac aggacacagc aatcaggta

721 gccaaaatta ccctatagtg cagaacatcc aggggcaaat ggtacatcag gccatatcac
781 ctagaacttt aaatgcattt gtaaaaatgt tagaagagaa ggcttcagc ccagaagtga
841 tacccatgtt ttccagcattt tcagaaggag ccacccacaa agatttaaac accatgctaa
5 901 acacagtggg gggacatcaa gcagccatgc aaatgttaaa agagaccatc aatgaggaag
961 ctgcagaatg ggatagatgt catccagtgc atgcagggcc tattgcacca ggccagatga
1021 gagaaccaag gggaaatgtac atagcaggaa ctactgtac ctttcaggaa caaataggat
1081 ggatgacaaa taatccacccat atccagtag gagaattttta taaaagatgg ataatccctgg
1141 gattaaataa aatagtaaga atgtatagcc ctaccagcat tctggacata agacaaggac
10 1201 caaaggaacc ctttagagac tatgttagacc ggttctataa aactctaaga gccgagcaag
1261 cttcacagga ggtaaaaat tggatgacag aaaccttgtt ggtccaaaat gcgaaacccag
1321 attgttaagac tattttaaaaa gcattggac cagcggctac actagaagaa atgatgacag
1381 catgtcaggg agtaggagga cccggccata aggcaagagt ttggctgaa gcaatgagcc
1441 aagtaacaaa ttcaagctacc ataatgtac agagaggcaaa tttaggaac caaagaaaga
15 1501 ttgttaagtg ttcaattgt ggcaaaaaggg ggcacacacgc cagaattgc agggccctta
1561 gaaaaaaggg ctgtggaaa tggaaagg aaggacacca aatgaaagat tgtactgaga
1621 gacaggctaa ttttttaggg aagatctggc cttcctacaa ggaaaggcca ggaaatttc
1681 ttcaagcagc accagagccaa acagccccac cagaagagag cttcaggctt gggtagaga
1741 caacaactcc ccctcagaag caggagccga tagacaagga actgtatcct ttaacttccc
20 1801 tcaggtcaacttggcaac gacccctcgta cacaataaag atagggggc aactaaagga
1861 agctcttataa gatacaggag cagatgatac agtattagaa gaaatgatgtt tgccaggaag
1921 atggaaacca aaaatgatag gggaaattgg aggttttacca aagtaagac agtatgatca
1981 gatactcata gaaatctgtg gacataaagc tataggtaca gtattagtag gacctacacc
2041 tgtcaacata attgaaagaa atctgttgc tcagattggt tgcaactttaa attttccat
25 2101 tagccctatt gagactgtac cagtaaaattt aaagccagga atggatggcc caaaaagttaa
2161 acaatggccaa ttgacagaag aaaaaataaa agcatttagta gaaatttgta cagagatgga
2221 aaaggaaggg aaaatttcaaa aatggggcc tggaaatcca tacaatactc cagttttgc
2281 catabagaaa aaagacagta ctaaatggag aaaatttagta gatgtcagag aacttaataa
2341 gagaactcaa gactctggg aagtcaattt aggaataccca catccgcag ggttaaaaaaa
30 2401 gaaaaaatca gtaacagtac tggatgtggg tggatgtcatat ttctcaggcc ctttagatga
2461 agacttcagg aagtatactg catttacccat accttagtata aacaatgaga caccaggat
2521 tagatatcg tacaatgtgc ttccacaggg atggaaaggg tcaaccagcaa tattccaaag
2581 tagcatgaca aaaatcttag agcccttttag aaaaacaaaat ccagacatag ttatctatca
2641 atacatggat gattgtatg taggatctga cttagaaataa gggcagcata gaacaaaaat
35 2701 agaggagctg agacaacatc tggatgtggg gggacttacc acaccagaca aaaaacatca
2761 gaaagaacctt ccattccctt ggatgggtt tgaactccat cctgataaaat ggacagtaca

2821 gcctatagt ctgccagaaa aagacagctg gactgtcaat gacatacaga agttatgtgg
2881 gaaattgaat tggcaagtc agatttaccc agggattaaa gtaaggcaat tatgtaaact
2941 ccttagagga accaaagcac taacagaagt aataccacta acagaagaag cagagctga
3001 actggcagaa aacagagaga ttctaaaaga accagtacat ggagtgtatt atgaccatc
3061 aaaagactta atagcagaaa tacagaagca gggcaaggc caatggacat atcaaattta
3121 tcaagagcca tttaaaaatc tgaaaacagg aaaatatgca agaatgaggg gtgcacac
3181 taatgtatgtaaacaattaa cagaggcagt gcaaaaaata accacagaaa gcatagtaat
3241 atggggaaag actcctaaat ttaaactgcc catacaaaaag gaaacatggg aaacatggtg
3301 gacagagtat tggcaagcca cctggattcc tgagtggag ttgttaata cccctccctt
3361 agtgaattaa tggtaccagt tagagaaaga acccatagta ggagcagaaa ccttctatgt
3421 agatgggcga gctaacaggg agactaaatt agggaaagca ggatatgtta ctaatagagg
3481 aagacaaaaaa gttgtcaccc taactgacac aacaaatcag aagactgagt tacaagcaat
3541 ttatctatgtt tgcaggatt cgggatttga agtaaacata gtaacagact cacaatatgc
3601 attaggaatc attcaagcac aaccagatca aagtgaatca gagtttagtca atcaaataat
3661 agagcagttttaaaaaaagg aaaaggctta tctggcatgg gtaccagcac acaaaggaaat
3721 tggagggaaat gaacaagtag ataaatttagt cagtgttgc atcagggaaag tactatTTT
3781 agatggaaata gataaggccc aagatgaaca tgagaaatcacatggaaat ggagagcaat
3841 ggctatgtat ttaacctgc cacctgtatg agcaaaagaa atagtagccat gctgtgataa
3901 atgtcagcta aaaggagaag ccatgcatgg acaagtagac tggatgtccag gaatatggca
3961 actagattgt acacatttag aaggaaaagt tafccgtta gcagttcatg tagccagtg
4021 atatatagaa gcagaagtta ttccagcaga aacagggcag gaaacagcat atttctttt
4081 aaaattagca ggaagatggc cagtaaaaaac aatacatact gacaatggca gcaatttcac
4141 cggtgcacg gtttagggccg cttgttgc ggcgggaaatc aaggcagaat ttggaaattcc
4201 ctacaatccc caaagtcaag gagttgttgc atctatgtaaatggaaatggaaat
4261 aggacaggta agagatcagg ctgaacatct taagacagca gtacaatgg cagtattcat
4321 ccacaattttt aaaagaaaag gggggattgg ggggtacagt gcagggaaa gaatgttgc
4381 cataatagca acagacatac aaactaaaga attacaaaaa caaattacaa aaattcaaaa
4441 tttcggttattacaggg acagcagaaa tccactttgg aaaggaccag caaagctcct
4501 ctggaaaggt gaagggccag tagtaataca agataatagt gacataaaaag tagtgc
4561 aagaaaagca aagatcatta gggattatgg aaaacagatg gcaggtgtatg attgtgtggc
4621 aagtagacag gatgaggatt agaacatggaaatgtttagt aaaacaccat atgtatgtt
4681 cagggaaagc tagggatgt tttatagac atcactatgaa aagccctcat ccaagaataa
4741 gttcagaagt acacatccc cttagggatg cttagattgtt aataacaaca tattgggtc
4801 tgcatacagg agaaagagac tggcattgg gtcagggagt ctccatagaa tggaggaaaa
4861 agagatatag cacacaagta gaccctgaac tagcagacca actaattcat ctgttattact

4921 ttgactgttt ttcaagactct gctataagaa aggccttatt aggacacata gttagcccta
4981 ggtgtgaata tcaaggcagg cataacaagg taggatctct acaaacttg gcactagcag
5041 cattaataac accaaaaaaag ataaaagccac ctttgcctag ttttacgaaa ctgacagagg
5101 atagatggaa caagccccag aagaccaagg gccacagagg gagccacaca atgaatggaa
5161 actagagctt ttagaggagc ttaagaatga agctgttaga cattttccta ggattttggct
5221 ccatggctta gggcaacata tctatgaaac ttatggggat acttgggcag gagttggaaagc
5281 cataataaga attctgcaac aactgctgtt tatccattt cagaattttt tgtcgacata
5341 gcagaatagg cgttactcga cagaggagag caagaaatgg agccagtaga tcctagacta
5401 gagccctgga agcatccagg aagtgcgcct aaaactgcctt gtaccaattt ctattgtaaa
5461 aagtgttgct ttcatggca agttgtttc ataacaaaag ctttaggcatt ctcctatggc
5521 aggaagaagc ggagacagcg acgaagagct catcagaaca gtcagactca tcaagcttct
5581 ctatcaaagc agtaagttagt acatgttaatg caacctatac caatagtagc aatagtagca
5641 ttagtagtag caataataat agcaatagtt gtgtggtcca tagtaatcat agaatatagg
5701 aaaatattaa gacaaaagaaa aatagacagg ttaattgata gactaataga aagagcagaa
5761 gacagtggca atgagagtga aggagaaaata tcagcacttg tggagatggg ggtggagatg
5821 gggcaccatg ctccctggga tggatgtatg ctgtgtgtc acagaaaaat tggggtcac
5881 agtctattat ggggtacctg tggaaagga agcaaccacc actcttattt gtgcattcaga
5941 tgctaaagca tatgatacag aggtacataa tggatggcc acacatgcct gtgtacccac
6001 agaccccaac ccacaagaag tagtattggtaatgtgaca gaaaatttta acatgtggaa
6061 aaatgacatg gtgaaacaga tgcattgagga tataatcagt ttatggatc aaagcctaaa
6121 gccatgtgtaaatataaccc cactctgtgt tagttaaag tgcactgatt tgaagaatga
6181 tactaataacc aatagtagta gccccggaaat gataatggag aaaggagaga taaaaaactg
6241 ctcttcaatc acacccaaac gcataagagg taaggcgcag aaagaatatg ctttttttta
6301 taaactgtatataatccaa tagataatga tactaccagg tataagtgtca caagtgtaa
6361 cacctcagtc attacacagg cctgtccaaa ggtatccctt gagccattc ccatacattt
6421 ttgtcccccg gctggtttgcgatctaaa atgtaataat aagacgttca atggaacagg
6481 accatgtaca aatgtcagca cagttacaatg tacacatggaa attaggccag tagtatcaac
6541 tcaactgctgtttaatggca gtcttagcaga agaagaggta gtaatttagat ctgtcaattt
6601 cacggacaat gctaaaaacca taatgtaca gctgaacaca tctgttagaaa ttaattgtac
6661 aagacccaaac aacaatacaa gaaaaagaat ccgtatccag agaggaccag ggagagcatt
6721 tggatcaata gaaaaatag gaaatatgag acaagcacat tgtaacatta gtagagcaaa
6781 atgaaataac actttaaaac agatagctag caaattaaga gaacaatttg gaaataataa
6841 aacaataatc ttaagcaat ctcaggagg ggacccagaa attgtaacgc acagttttaa
6901 ttgtggaggg gaaattttctt actgttaattc aacacaactg ttaatagta ctgggtttaa
6961 tagtacttgg agtactgaag ggtcaataa cactgaagga agtgacacaa tcaccctccc

7021 atgcagaata aaacaaatta taaacatgtg gcagaaaagta ggaaaagcaa tgtatcccc
7081 tcccatcagt ggacaaatta gatgtcatc aaatattaca gggctgctat taacaagaga
7141 tggtgtaat agcaacaatg agtccgagat cttcagacct ggaggaggag atatgaggga
5 7201 caattggaga agtgaattat ataaatataa agtagtaaaa attgaaccat taggagtagc
7261 acccacaaag gcaaagagaa gagtggtgca gagagaaaaa agagcagtgg gaataggagc
7321 ttgtccctt gggttcttgg gaggcagg aagcactatg ggcgcagcc caatgacgt
7381 gacggtacag gccagacaat tattgtctgg tatagtgcag cagcagaaca atttgctgag
7441 ggctattgag ggcgaacagc atctgttgc actcacagtc tggggcatca agcagctcca
10 7501 ggcaagaatc ctggctgtgg aaagataacct aaaggatcaa cagctccctgg ggatttgggg
7561 ttgctcttggaa aaactcattt gcaccactgc tgccttgg aatgcttagtt ggagtaataa
7621 atctctggaa cagatttggaa atcacacgc ctggatggag tgggacagag aaattaacaa
7681 ttacacaagc ttaatacact ccttaatfta agaatcgca aaccagcaag aaaagaatga
7741 acaagaatta ttgaaattag ataaatgggc aagtttgg aattggtttta acataacaaa
15 7801 ttggctgtgg tatataaaat tattcataat gatagtagga ggcttggtag gtttaagaat
7861 agttttgtctactttctta tagtgaatag agttaggcag ggatattcac cattatcggt
7921 tcagacccac ctcccaaccc cgaggggacc cgacaggccc gaaggaatag aagaagaagg
7981 tggagagaga gacagagaca gatccattcg attagtgaac ggatcccttgg cacttacgt
8041 ggacgatctg cggagcctgt gccttctcag ctaccaccgc ttgagagact tactcttgat
20 8101 tgtaacgagg atttggAAC ttctggacg caggggtgg gaagccctca aatattggtg
8161 gaatctccta cagtatttggaa gtcaggaact aaagaatagt gctttagct tgctcaatgc
8221 cacagccata gcagtagctg agggacaga tagggtata gaagtagtac aaggagctg
8281 tagagctatt cgccacatac cttagaagaat aagacagggc ttggaaagga ttttgctata
8341 agatgggtgg caagtggtaaaaatgt tgattggatg gcctactgt aggaaagaa
8401 tgagacgagc tgagccagca gcagataggg tggagcagc atctcgagac ctggaaaaac
25 8461 atggagcaat cacaagttagc aatacagcag ctaccaatgc tgcttgc tggctagaag
8521 cacaagagga ggaggagggtg gttttccag tcacacctca ggtacccttta agaccaatga
8581 cttacaaggc agctgttagat cttagccact tttaaaaaga aaaggggggc ctggaaaggc
8641 taattcactc ccaaagaaga caagatatcc ttgatctgtg gatctaccac acacaaggct
30 8701 acttccctga tttagcagaac tacacaccag ggccagggtt cagatatcca ctgacccttg
8761 gatggtgtaa caagcttagta ccagttgagc cagataagat agaagaggcc aataaaggag
8821 agaacaccag ctgttacac cctgtgagcc tgcatggat ggtaccccg gagagagaag
8881 tgtagagtg gaggttgc acgcgcctag catttcatca cgtggcccgagactgc
8941 cggagactt caagaactgc tgacatcgag ctgttacaa gggacttcc gctggggact
9001 ttccaggag gctggccctg ggccggactg gggagtgccg agccctcaga tcctgcata
35 9061 aagcagctgc ttttgcctg tactgggtct ctctggtag accagatctg agcctggag

9121 ctctctggct aactaggaa cccactgctt aagcctcaat aaagcttgcc ttgagtgctt
 9181 c (SEQ ID NO: 14)

5 Initial Specific Target Motifs:

- (1) Trans-activation response region/Tat protein binding site - TAR RNA - nts 1 - 60
 "Minimal" TAR RNA element
 5' GGCAGAUCUGAGCCUGGGAGCUCUCUGCC 3' (SEQ ID NO: 15)
- 10 (2) Gag/Pol Frameshifting Site - "Minimal" frameshifting element
 5' UUUUUUAGGGAAGAACUGGCCUUCCUACAAGGGAAGGCCAGG
 GAAUUUUUCUU 3' (SEQ ID NO: 16)

5.7. Hepatitis C Virus ("HCV" - Genotypes 1a & 1b)

15 GenBank Accession # NC_001433:

1 ttggggcga cactccacca tagatcactc ccctgtgagg aactactgtc ttacgcaga
 61 aagcgtctag ccatggcggt agtatgagtg ttgtcagcc tccaggaccc cccctcccg
 121 gagagccata gtggctcgcc gaaccggta gtacaccgga attgccagga cgaccgggtc
 181 ctttcttggta tcaaccggct caatgcctgg agatttggcc gtgcggccgc gagactgcta
 20 241 gcccggatgt gtgggtcgcc gaaaggccctt gtggactgc ctgatagggt gcttgcgagt
 301 gccccggag gtctcgtaga ccgtgcatca tgagcacaaa tcctaaacct caaagaaaaa
 361 ccaaacgtaa caccaaccgc cgccccacagg acgttaagtt cccggggcgtt ggtcagatcg
 421 ttggggaggat ttacctgttg cggcgccagg gccccagggtt ggggtgcgc gcgacttagga
 481 agactcccgaa ggggtcgcaat cctcgtggaa ggcgacaacc tatccccaaag gctcgccggc
 541 ccggggtag gacctggct cagccgggtt acccttggcc cctctatggc aacgagggtt
 601 tggggggcaggatggctc ctgtcacccctt gtggctctcg gcctagttgg ggcggccacag
 661 acccccccggcg taggtcgctt aattttggta aggtcatcgta tacccttaca tgcggcttcg
 721 ccgacccat ggggtacatt ccgttgcgtc ggcggccctt agggggcgctt gccaggccc
 781 tggcacatgg tggccgggtt ctggaggacg ggcgtgaacta tgcaacaggaa aatctgccc
 841 gttgtctttt ctctatcttc ctcttagttt tgctgtcttg ttggaccatc ccagcttccg
 901 cttacgagggtt ggcgaacgtt tccggatattt accatgtcac gaaacgtactgc tccaaactcaa
 961 gtatgtgtat tgaggcagcg gacatgtatca tgcacaccccc cgggtgcgtt cccgtcgcc
 1021 gggagagtaa ttctccgtt tgctggtagt ccgtcactcc cacgtcgccg gccaggaaca
 1081 gcagcatccc caccacgaca atacgacgcc acgtcgattt gctcggtgg gccggctgtc
 1141 tctgtccgc tatgtacgtt gggatctctt gggatccgtt ttctcgcc tccctcgatgt
 1201 tcacccatcc acctcgccgg tatgagacgg tacaagattt caatgtcata atctatcccc

1261 gccacgtatc aggtcaccgc atggcttggg atatgtatgt gaactggtca cctacaacgg
 1321 ccctagtgtt atcgcagctt ctccggatcc cacaagccgt cgtggacatg gtggcgggggg
 1381 cccactgggg tgccttagcg ggccttgccct actattccat ggtggggaaac tgggctaaagg
 5 1441 tcgttattgt gatgtactc ttgtctggcg ttgacggca caccacgtg acagggggaa
 1501 gggtagcctc cagcacccag agcctcggt cctggcttc acaaggccca ttcagaaaa
 1561 tccaactcgta gaacaccaac ggcagctggc acatcaacag gaccgctctg aattgcaatg
 1621 actccctcca aactgggttc attgtcgcc tgcgtacgc acacagggttc aacgctccg
 1681 ggtgcccaga ggcacatggct agctgcccgc ccatcgatga gttcgctcg ggggtgggtc
 10 1741 ccatcactca tgcgtatgcct gagagctgg accagaggcc atattgtgg cactacgccc
 1801 ctgcaccgtg cgggatcggt cctgcgtcgc aggtgtgtgg tccagtgtat tgcttcactc
 1861 cgagccctgt tgcgtgggg acgaccgatc gttcggcgc tcctacgtat agctgggggg
 1921 agaatgagac agacgtcgct ctacttagca acacgcccgc gcctcaaggc aactggtttgc
 1981 ggtgcacgtg gatgaacagc actgggtca ccaagacgtg cggggccctt ccgtcaaca
 15 2041 tcgggggggt cggcaacaac accttggtct gccccacgga ttgttccgg aagcaccccg
 2101 aggcacatca cacaaggatgt ggctcggggc cctgggtgac acccagggtgc atgggtgact
 2161 acccatacag gctctggcac taccctgca ctgttaacit taccgtctt aaggtcagga
 2221 tgcgtatgttggg gggcgtggag cacaggctca atgcgtcatg caattggact cgaggagagc
 2281 gctgtgacatgggggacaggatggcgtacgg aactcagccgc gctgtgtcg tctacaacag
 20 2341 agtggcagat actgcccgt tccttcacca ccctaccggc cctgtccact ggcttgatcc
 2401 atcttcacccg gaacatcggt gacgtcaat acctgtacgg tatagggtcg gcaatgttct
 2461 ccittgcaat caaatgggag tataccgttgc tgcgttccct tccttcggcgc gacgcgcgc
 25 2521 tctgtgcctg ctgtggatg atgcgtctga tagccaggc tgaggccacc ttagagaacc
 2581 tgggtgtccct caatgcggcg tctgtggccg gagcgtatggc ccttcttc tccctgtgt
 2641 tcctctgcgc cgcctggatc atcaaaggca ggctggccc tggggcggca tatgtctct
 2701 atggcgatg gccgttgctc ctgtcttgc tggcattacc accacgagct tatgcatgg
 2761 accgagagat ggctgcgtatggcgtatggcgtatggcgtatggcgtatggcgtatggcgtatggc
 2821 tgcgtatggcgtatggcgtatggcgtatggcgtatggcgtatggcgtatggcgtatggcgtatggc
 2881 ccagagccga ggcgcacttgc caagtgtggg tccccctct caatgttgcgg ggaggccgc
 30 2941 atgcgtatcat cctccatata tcgtcggtcc atccagatgt aatcttgcgtatggcgtatggc
 3001 tcctgtcgccatactcggtt cgcgtatggcgtatggcgtatggcgtatggcgtatggcgtatggc
 3061 atcttgcgtatggcgtatggcgtatggcgtatggcgtatggcgtatggcgtatggcgtatggc
 3121 gcccactatgtt ccaatggcc ttcatgtatggcgtatggcgtatggcgtatggcgtatggc
 3181 accatcttac tccactgcgg gattggccc acgcggccctt acgagacccgtt gcggtggc
 3241 tagagccgtt cgttctctt gacatggaga ctaaactcat caccgtgggg gcaacacccg
 3301 cggcgtgtgg ggacatcatc tcgggtctac cagtctccgc ccgaagggggg aaggagatac

5461 gtgcttcaca cctcccttac atcgagcaag gaatgcagct cgccgagcaa ttcaaacaga
5521 aggcgcctcg attgctgcaa acagccacca agcaagcgga ggctgctgct cccgtgg
5581 agtccaagtg gcgagccctt gaggtttctt gggcgaaaca catgtggAAC ttcatcagcg
5641 ggatacaga cttggcaggc ctatccactc tgcctggaaa ccccgcgata gcatcattga
5701 tggctttac agcctctatc accagccgc tcaccaccca aaataccctc ctgttaaca
5761 tcttgggggg atgggtggct gcccaactcg ctccccccag cgctgctcg gctttcg
5821 ggcgcggcat tgccggtgcg gccgtggca gcataggctc cgggaaggta cttgtggaca
5881 ttctggcggg ctatggggcg ggggtggctg ggcactcg ggccttaag gtcatgagcg
5941 gcgagatgcc ctccactgag gatctggta attactccc tgccatcctt tctctggcg
6001 ccctgggtgt cggggcgtg tgccgcgca tactgcgtcg gcacgtggc cggggagagg
6061 gggctgtca gtggatgaac cggctgatag cgttcgttc gccccggtaac cacgtctccc
6121 ccacgcacta tggcccgag agcgcacgcg cggcgcgtgt tactcagatc ctctccagcc
6181 ttaccatcac tcagttgctg aagaggcttc atcagtggat taatgaggac tgctccacgc
6241 ctgtccgg ctgcgtggca aaggatgtt gggactggat atgcacggtg ttgagtgact
6301 tcaagacttg gctccagtc aagctctgc cgcgggttacc gggactccct ttccgtcat
6361 gccaacgcgg gtacaaggga gtctggcggg gggatggcat catgcaaacc acctgc
6421 gtggagcaca gatcaccggaa catgtaaaa atggctcat gaggattgtt gggccaaaa
6481 cctgcagcaa cacgtggcat ggaacattcc ccatcaacgc atacaccacg ggc
6541 cgcctcccc aggcgcgaac tattccaggc cgctgtggcg ggtggctgtc gaggagtag
6601 tggagggtac gcgggtgggg gatttccact acgtgcacggg catgaccact gacaacgtga
6661 aatgcccattc ccagggttcca gcccctgaat tttcacggaa ggtggatggaa gtacgg
6721 acaggatgc tccagtgtaaa acacccatc tacgagagga ggtcgtaatc caggcgg
6781 tcaaccagta cctggcggg tcacagtc catgtggcc cgaaccggat gtggcagtg
6841 tcactccat gtcacccgac ccctctata ttacagcaga gacggccaag cgtaggctgg
6901 ccagggggtc tccccctcc ttggccagct cttagctcg ccagttgtct ggc
6961 tgaaggcgcac atgtactacc catcatgact cccggacgc tgacccatc gaggccaa
7021 tcctgtggcg gcaggagatg ggccggaaaca tcacccgtgt ggagtcagaa aataagg
7081 taatccatc ctcttcgtatccatccatccatccatccatccatccatccatccatccatccat
7141 cggcggagat cctgcgaaaa cccaggaaat tccccccagc gttggccata tgggcacgc
7201 cggattacaa ccctccactg cttagtgcctt ggaaggaccc ggactacgtc cccccc
7261 tacacgggtg cccttgcata tctaccagg ccccccata accacccatccatccatccat
7321 ggacgggtgt cctgcacagag tccaccgtgt ctctgcctt ggcggacgc gctactaaga
7381 ccttggcag ctccgggtcg tcggccgttg acagcggcac ggcgcactggc cctcc
7441 aggcctccga cgacggcgcac aaaggatcg acgttgatgc gtactccatc atgc
7501 tcgaggggaga gccaggggac cccgacccatca ggcacgggtc ttggtctacc gtgagcgggg

5 7561 aagctgggtga ggacgtcg tcgtgtctaa tgcctatac atggacagg gccttgatca
7621 cggcatgcgc tgcggaggag agcaagtgc ccatcaatcc gttgacaac tctttgtgc
7681 gtcaccacag tatggctac tccacaacat ctgcagcgc aagtctgccc cagaagaagg
7741 tcacccatgtc cagactgcaaa gtcctggacg accactaccg ggacgtgctc aaggagatga
7801 aggcaaggc gtccacagtt aaggcttaggc ttctatctat agaggagggc tgcaaaactga
7861 cgccccaca ttggccaaa tccaaatitg gtcacggggc gaaggacgctc cggagcctat
7921 ccagcaggc cgtcaaccac atccgctccg tgtggagga cttgctggaa gacactgaaa
7981 caccaattgtc taccaccatc atggcaaaaa atgaggttt ctgcgtccaa ccagagaaag
10 8041 gaggccgcaaa gccagctcg cttatcgat tccagaccc ggggtacgt gtatgcgaga
8101 agatggccctt acgacgtg gtcctccccc ttccctaggc cgtgatgggc ccctcatacg
8161 gattccagta ctctctggg cagcgggtcg agttcctggt gaataccctgg aaatcaaaga
8221 aatgcctat gggctctca tatgacaccc gctgttgc ctcaacggc actgagaatg
8281 acatccgtac tgaggaatca atttaccaat gtgtgactt gcccccgaa gccaggcagg
15 8341 ccataaggc gtcacagag cggctttagt tcgggggtcc cctgactaat tcgaaggggc
8401 agaactgcgg ttatgcgg tgccgcgcaaa gtggcgtgtc gacgactagc tgccgcaaca
8461 ccctcacatg ttacttgaag gccactgcgg cctgtcgagc tgcaaaagctc caggactgca
8521 cgatgctgt gaacggagac gacccgtcg ttatctgtga gatgtcgggaa acccaggagg
8581 atgcggcggc cctacgagcc ttacggagg ctatgactag gtattccgccc cccccgggg
20 8641 accccggccca accagaatac gacttggagc tgataacgtc atgctccccc aatgtgtcg
8701 tcgcgcacga tgcattccggc aaaagggtgt actacccac ccgtgacccc accccccccc
8761 tcgcacgggc tgcgtggag acagtttagac acactccagt caactccctgg ctaggcaata
8821 tcatcatgtc tgcggccacc ctatggcga ggtgttctt gatgactcat ttcttctctaa
8881 tccttctac tcaggagcaa ctggaaaaag ccctggattt tcagatctac gggccctgtt
25 8941 actccattgtc gcccattgtac ctacccatc tattgtacg actccatggt cttagcgcatt
9001 ttctacttcca cagttactctt ccaggtgaga tcaatagggtt ggcttcatgc ctcaggaaac
9061 ttggggtacc gccttgcga gtctggagac atcggggccag aagtgtccgc gctaagctac
9121 tgcgtggggc gggagggttgc gcaaggatctt ctcaactgg gcagtaaaaga
9181 ccaagcttac actcacttcca atccggctg cgtccctggt agactgtcc ggctgggttgc
9241 ttgtgttcaacggggga gacatatac acagccgtc tcgtgcccga ccccggttgg
30 9301 tcatgtgtc cttacttcca ctgtgttagt gggtaggcattt ctacccatgtc cccaaacgggt
9361 gaacggggag ctaaccactc caggccataa ggccattccc tttttttttt ttc (SEQ ID NO: 17)

General Target Region:

35 5' Untranslated Region - nts 1 - 328 - Internal Ribosome Entry Site (IRES):

5'UUGGGGGCGACACUCCACCAUAGAUCACUCCCCUGUGAGGAACUACUGUCUU
 CACGCAGAAAGCGUCUAGCCAUGGCGUUAGUAUGAGUGUUGUGCAGCCUCCA
 GGACCCCCCUCCCGGGAGAGCCAUGUGGUCUGCGGAACCGGUGAGUACACC
 5'GGAAUUGCCAGGACGACCGGGUCCUUUCUJUGGAUCAACCCGCUAAUGCUGG
 AGAUUUGGGCGUGCCCCCGCAGACUGCUAGCCAGUAGUGUUGGGUCGCGA
 AAGGCCUUGUGGUACUGCCUGAUAGGGUGCUUGCGAGUGCCCCGGGAGGUCU
 CGUAGACCGUGCAU3' (SEQ ID NO: 18)

10 Initial Specific Target Motifs:

- (1) Subdomain IIIc within HCV IRES - nts 213 - 226
 5'AUUUGGGCGUGCCC3' (SEQ ID NO: 19)
- (2) Subdomain IIId within HCV IRES - nts 241-267
 5'GCCGAGUAGUGUUGGGUCGCGAAAGGC3' (SEQ ID NO: 20)

15

5.8. Ribonuclease P RNA ("RNaseP")

GenBank Accession #s

X15624 Homo sapiens RNaseP H1 RNA:

1 atggcggag ggaagctcat cagtggggcc acgagctgag tgcgtctgt cactccactc
 61 ccatgtccct tgggaaggtc tgagactagg gccagaggcg gcccataacag ggctctccct
 121 gagttcagg gaggtgagtt cccagagaac ggggctcccg gcgaggtag actggcagg
 181 agatgccgtg gaccccgccc ttccgggagg ggcccgccgg atgcctccct tgccggagct
 241 tggAACAGAC tcacggccag cgaagtgagt tcaatggctg aggtgaggta ccccgccagg
 301 gacctcataa cccaaatcag accactctcc tccgcccatt (SEQ ID NO: 21)

25

U64885 Staphylococcus aureus RNaseP (rrnB) RNA:

1 gaggaaagtc cgggctaca cagtcgtaga tgattgtgt gttcgtgttt gatgaaacaa
 61 taaaatcaagg cattaatttg acggcaatga aatatctaa gtccttcgtat atggatagag
 121 taatttggaaa gtgccacagt gacgttagtt ttatagaaat ataaaagggtg gaacgcggta
 181 aacccctcga gtgagcaatc caaatttggt aggagcactt gtttaacgga attcaacgt
 241 taaacgagac acacttcgctg aaatgaagtgt gtgttagacag atggttatca cctgagtacc
 301 agtgtgacta gtgcacgtga tgagtacgtat ggaacacagaac gccggcttat (SEQ ID NO: 22)

35

M17569 Escherichia coli RNA component (M1 RNA) of ribonuclease P (rnpB) gene:

1 gaagctgacc agacagtgcgc cgcttcgtcg tcgtcctttt cgggggagac gggcggaggg

61 gagggaaagtc cgggctccat agggcagggt gccaggtAAC gcctgggggg gaaacccacg
 121 accagtgcaa cagagagcaa accgcccgtg gcccgcgcaa gggggatcag gtaagggtga
 181 aagggtgcgg taagagcgca ccgcgcggct ggtAACAGTC cgtggcacgg taaactccac
 5 241 cggagcaag gcAACatagg ggttcatAGT gtacggcccg tactgaACCC gggtaggctg
 301 cttgagccag tgagcgattg ctggcctaga tgaatgactg tccacgacag aacccggctt
 361 atcggtcagt ttaccc (SEQ ID NO: 23)

Z70692 Mycobacterium tuberculosis RNaseP (rnpB) RNA:
 10 1 ccaccgtta cgtatccgtcc gaccatggcc ccacaatagg gcccggaga cccggcgta
 61 gtggtggcg gcacggtcag taacgtctgc gcaacacggg gttgactgac gggcaatatc
 121 ggctccatag cgtccggccgc ggatacagta aaggagcatt ctgtgacgga aaagacgccc
 181 gacgacgtct tcaaacttgc caaggacgag aaggcgtaa atgtcgacgt cccgttctgt
 241 gacccgtcg gcatcatgca gcaacttcacg attccggctt cggcccttga caagagcgtg
 15 301 ttgacgacg gcttggccctt tgacggctcg tcgattcgcg gttccagtc gatccacgaa
 361 tccgacatgt tgcttcttcc cgatcccgag acggcgcgca tcgaccggctt cccgcggcc
 421 aagacgctga atatcaactt ctttgtcac gaccgttca ccctggagcc gtactcccg
 481 gacccgacca acatcgcccg caaggccgag aactacctgta tcagcactgg catcgccgac
 541 accgcatact tcggcgccga ggccgagttc tacatttgcg attcggtgag ctgcactcg
 20 601 cggccaaacg gctccctcta cgagggtggac gccatctgg ggtggtgaa caccggcg
 661 cgcaccgagg cggacggcg tcccaaccgg ggctacaagg tccggccacaa gggccggat
 721 ttcccaacttgg ccccaacga ccaatacgctc gacctgcgcg acaagatgtt gaccaacctg
 781 atcaactccg gtttcatccg ggagaaggc caccacgagg tggcagcgg cggacaggcc
 841 gagatcaactt accagttcaa ttgcgtctg cacggcccg acgacatgca gttgtacaag
 901 tacatcatca agaacaccgc ctggcagaac ggcaaaacgg tcacgttcat gccaaaggcg
 961 ctgttggcg acaacgggtc cggcatgcac tgcgttcatcgt cgctgtggaa ggacggggcc
 1021 ccgttgcgtt acgacgagac gggttatgcc ggttgcgcg acacggcccg tcattacatc
 1081 ggcggcctgt tacaccacgc ggcgtcgctg ctggcccttca ccaacccgac ggtgaactcc
 1141 tacaagcggc tggttcccggtt acgaggcc cggatcaacc tggtctatag ccagcgcaac
 1201 cggtcggcat gctgtcgcat cccgatcacc ggcagcaacc cgaaggccaa ggggtggag
 1261 ttccgaagcc cggactcgctc gggcaacccgg tatctggcg tctcgccat gctgtggca
 1321 ggcgttggacgtt acatcaagaa caagatcgag cggcaggcgc cggcgttacaa ggatcttac
 1381 gagctgcgcg cggaaagaggc cggcgttac cggcagactc cggcgttacgt gtcagatgt
 1441 atcggccgtc tcgaggccga ccacgaatac ctcaccgaag gagggtgtt cacaacgcac
 1501 ctgtatcgaga cgtggatcag ttcaagcgca gaaaacgaga tcgagccgtt caacatccgg
 1561 cggcatccctt acgaattcgcc gctgttactac gacgtttaag gactcttcgc agtccgggtg

1621 tagagggagc ggcgtgtcgt tgccagggcg ggcgtcgagg ttttcgtatg ggtgacggtg
1681 gccggcaacg ggcgcggac caccgctcgtc aagagcccgtaaagaacgt tcaaggacgt
1741 ttcagccggg tgccacaacc cgcttggcaa tcatctcccg accgcccggc ggggttgcit
5 1801 tcacatgcgc cggaaactcaa gccacgtcgt cgcccaggcg tgcgtcgcg gcccggtcag
1861 gttaaatgtc ggggattcgt cgtgcggcg ggcgtccacg ctgaccaacg gggcagtcaa
1921 ctcccaaca ctttgcgcac taccgcctt gcccggcg tcaccctgt tagttgtcc
1981 aggaattccc caccgtcg tttcgccag cggccgcga cggcgaccgc attgagctgg
2041 cggccgggtc cggcgacgtg tgcgtgggc ttggccgcga ccaacaccag cgcgttgcgg
10 2101 gcccgggtgg cggtcagcca ggcctgacgg agcagctcca cgtcggtgc gggaaaccaga
2161 tcggcgccg cgatgacatc cagggattgc agcgtcgagg tgcgtgcag ggcgggaacc
2221 tggtgcgcat gctgttagctg cagcaactgc acggtccatt cgtatgcggc cagtccggc
2281 cggcccagt tggtgtgtgt tttggggctcg gcaccgcgc gcaaccgcgc ggactcgata
2341 cgggccttga tgcggcgaat ctgcgcacc gagtcagcgg acacaccgtc gggggatac
15 2401 cgcgtttgtt cgtaccatccg taggaatcgc tgacccaaact cggcatcgcc ggcaaccgc
2461 tgcgcgtta gcagggcctg gatctccat ggctgtgcgc actgctcgta gatgcggcg
2521 taggacccca gggtgcggac cagcggaccg ttgcggccct cgggtcgcaa attggcgtcg
2581 agctccagcg gggatgcac gctgggtgtc cccagcagcg cccgaaccccg ctgcgcgtac
2641 gatgtcgacc atttacccgc cctgtcatcg tcgacgcgg tggccggctc acagacgaac
20 2701 atcacgtcg catccgaccc gtagccaaac tcggcaccac ccagccgacc catgcgtatg
2761 accgcgtatgg ccgcggggc gcgatgtcg tcgggaaggc tggccggat catgacgtcc
2821 agcgcggcct gcagcaccgc caccacacc gacgtcaacg cccggcacac ctgcgtgacc
2881 tcgagcaggc cgaggcgtc cggcgaaccg atgcgggca gcttcgacg acgcagcgtg
2941 cgcgcggcg cgtatggccct tcgggtgtc gggtagcggc tcgccgaggc gatcagcgcc
25 3001 cgagccacgg cgccgggtc ggtctcgagc agcttcgggc ccgcaggccc gtcctgtac
3061 tgctggatga cccgcggcgc gcgatcaac agatccggca catacgccga ggtacccaaag
3121 acatgcgtatga ggcgttggc caccgcggc tttccgcga gctgtggccag gtaccagctt
3181 tcgggtggca gcgcctact gagccgcgg taggccagca gtccgcgtc gggatcgggg
3241 gcatacgaca tccagtccag cagcgtggc agcagcaccg actgcacccg tccgcgcgg
3301 cccgtttat tgaccaacgc cgacatgtgt tcaacgcgg tgcgttgcgtcc ctgttagccc
3361 agcgcggcca gcccggccc cggccctcc aacgtcatgc cgtggcgat ctccaacccg
3421 gtcggccga tcgttccag cagcgttga tagaagagtt tggtgtttaa ctgcacacc
3481 cgcacgttct gtttcttgcg ttcctccgc agcaccggc cgcacatgtt tcggccatcg
3541 ggcggatgt gggccgcgc cggccagccag cgcactgcct cctcgcttc gggatcggga
3601 agcaggtggg tgcgttgcg cgcgttgcac tgcgttgcgt gtcgagcag cctgaggaaac
35 3661 tcatacgacg cggcatgtt cgcgcgtcc tcacgcggc tgcgttgcgtcc ttgcggcaac

3721 gcccatacg cgtccaccgt ggacgccacc cgtaacgact cgtcgtacg ggcataacc
3781 agtcgcagta gctgtacggc gaactccacg tcgcgcatac cgccgtgcc gagtttggc
3841 tcgcggccgc ggacatcgcc gggcaccaggc tgcacccatc ggcgcgcac ggctgcacc
5 3901 tcgaccacaa agtctcgcg ctgcaggct cgcacacca tcggcatcaa ggccgtcagg
3961 taacgcgtcgc caagttccgc gtcgcacacg actggccgtg ctttcagcaa cgctgaaac
4021 tcccaggctt tggcccgagcg ctggtagtag gcgatgtgcg actcgagcgt acggaccaggc
4081 tcccggtgc gccctccgg acgcaggcg gcgtccacct cgaaaaaggc cgccgaggcc
4141 accgcacatca tctcgctggc cacgcgcgcg ttgcgcgggt cggagcgctc ggcaacgaat
10 4201 atgacatcgac cgtcgctgac gtatcgttgc acgcacttgc catcgcgatg
4261 accggccaggc gcggtggcgg gtgcgtccgc cacacgcgtc cctcgccac ggcgcaggcc
4321 gcccgcagag cggcgccgc ggcgtccgcg aggcggtcg ccaccacggt gaatggcagc
4381 accgggttcgt cctcgaccgt cgccgcaggc tgcagagcgcc ccaggatttgc acgtatcg
4441 cggtaactggg ttgcataatcg gtgcacgcgc gaggccggca taccctccga ttctcgacg
15 4501 cactcgacga acgaccgctg cagctggta tggacggca gtgtgaccct gccccgcagc
4561 aatttccagg actcggtatg ggcgaccagg tgatcgccca acgcgcgtcg cgagcccgac
4621 accgagaaca gcccgcgcg cagactgcgt tcgcgcagca gagccgcgtt gagctcgcc
4681 catccgggtt ctggattctc cgacagccgg atcaaggcgcc gacgcgcggc atcggcggtcc
4741 ggagcgcgtg acagcgcacca cagcaggctg acgtgcgcct gatccctcgatc cgcgcgcac
20 4801 cccagctgag ccagacgctc accagcagggg gggtaacta atccgagccg gccaacgcgt
4861 ggcaacttcg gccgcgtcggtt ggcgatgttgc acgcgttgc gcaaggcgcc
4921 tcggcggtcg atcaaccggt agatctggc tacagcgcaca ggttagtgcg cagctcgat
4981 ggcgtgacgt ggctgcggta gttcgccac tccgtgcgt tggcgccaa gaaaaagtca
5041 aaaacgtgtt ccccaaggc ctccgcgcac agtgcggagg cctccatggc ggcgcgcac
25 5101 ctatccaaac tggacggcaa ttctcggtac cccatcgctc gggtgttgcgg
5161 tccccatacg tgcgtcgcc ctgcggccccc agcacgtaa ccttctctac accccgcata
5221 cccgcggccca gcagcacggc gaatgtcaga tagggattgc acgcgcatac agggctgcgt
5281 acitcgaccc gcccgcacga ggtttgtgc ggcgtgtaca tcggcaccgg cactaggcg
5341 gatccgggtt cggcccccac caacgcggcc gtggcgccctt cccgcgcgtt caccagccgc
30 5401 ttttaagatgtt gacccactg atttgttgc acgcgtatcg cgcgcgttgc ctccaggatc
5461 cggcgatgtt acgatttacc cacttcgcac agtgcgcgtc gatcatcgc gctgtggaaac
5521 ggcgttgcacat caccctcgaa caggctcatg tgggtgtca tcggcgcgc ggggtgtgg
5581 ccgaatggct tggcatgaa cgacgcccgg ggcgcgcctt ccagcgcgcac ttctttgtat
5641 acgtacggc aggtcatc acgtgtcagcc atgcacagag cgtcgccaaa cgcgcaggcgt
35 5701 atctccgtgtt ggcgggtgc gcctcgatc tggcgactt ccaccgagat gcccgtat
5761 tccaggccat cgtacgcgtg gcccggaaat ttcaaggcg agtgcgtgcac cgcgttgc

5 5821 aaatagccgg cggtgtcgac cgggacgggc accgaccctgt cctcgggtcc gggcttgagc
5881 aggaagaact cgatttcggg atgcacgtag caggagaagc cgagttcgcc ggccttcgtc
5941 agctggccgc gcaacacgtg cgcgggtcc gcccacgacg gcgagccgtc cggcatggtg
6001 atgtcgcaaa acatccgcgc tgagtggtgg tggccggaac tggtgccca gggcagcacc
6061 tggaaggtcg acgggtccgg gtgcgccacc gtatcggtt ccgagacccg cgaaagccc
6121 tcgatcgagg atccgtcgaa gccgatgcct tcctcgaagg cgcctcgag ttggctggg
6181 gcgtatggcga ccgacttgag gaaaccgagc acgtctgtga accacagccg gacgaagcgg
6241 atgtcgctt ctccagggt acgaagaacg aattccttcgttgcgtccat acctcgaaca
10 6301 gtatgcactg tctgttaaaa ccgtgttacc gatgcccggc cagaagcggtt gcggggcggc
6361 ccgcaagggg agtgcgcgtt gagitcaggg cgccacccgc agactctgtc gggcaaggt
6421 cccgtcgaga aaatagtgcata caccgcaga gtccacacac tggtgccat cgaacacccg
6481 agtgtgttgg gtgcgtcgaa aggtgatcag cggtgccccc agctggcggg ccaggctac
6541 cccggactga tacggagtgg ccgggtcggtt ggtggggac accacgacga cttggccagc
6601 cccggccggc gccgcgggtt gcggcgtcgaa cttggccggc accggccaca ggcgcacag
6661 atcgccccggg gcggatccgg tgaactgcctt gtagctaagg aacggggcga cctgacggat
6721 ccgttggtcg cggccaccc aggccgttgg atcggccgtt gttggccat cgcacccg
6781 gaccgcgtt aacgcgttgcgtt gtagtgccttgcgtt tctgcataccc ggccgtcata
6841 gtcgtcgca agcaccagca agtcgcggc gtcgtcgcc cgctgcagcc ccagcagacc
20 6901 actggtcagg tacttccagc gtcgttgcgtt gtagcgcgtt ttgtgggtc cggcgtcgc
6961 gtcggcgttag ctcaaggccac gtggatccga cttggccatcc ggcttctgc ctaggggtc
7021 aaccagggcg tggtagcggt tgaccactgt ggccgagtcg gtggccagag ggcaggccgg
7081 cgagccccggc cagtcggcgg cgtgtcatt gaaagcggtc tgaaatcccg ccattttggct
7141 gatgtttcc tcgattgggc taacggctgg atcgatagcg cggcgttgcgtt ccatcgcccg
25 7201 cacatgagta ccgaaccgtt ccaggtaagc ggtggccaaac tcgggtccgtt agctgttatcc
7261 gaggttagttt atctgtatgtt cacctaaccgc ttggcgaacc atgtccatgtt cccgtgcgtac
7321 ggacgcggta ccgtatattgg ccaagaagctt gaagcccatc cggtaacac acgtcctggc
7381 caactggccgg tagacctgtt cgacgtgggtt gacaccggcc ggactgttagt cggccatcggt
7441 atcgccggc tacgcgtcgaa actcgccgtc ggtgcgttgcgtt ccacacccg gggcgttgcgtt
7501 gccgaccctt ctccgggtcgaa agccccccagg gtcgttgcgtt gggagaatgtt cgggttcggc
7561 gatcgccgggtt gccatagcggtt cgaccatgtc gaccggccgtt gccccgggtt cccaggattt
7621 gaccaggcgtt gctccgttgcgtt cggccgggttgcgtt cggatcaccgtt ccaacttcgtt
7681 ttgtgtccca cccgggttgcgtt cgtgtcgac ggggacggac accgtcgccgtt agcgtcgact
7741 gcgtatccgtt cttggcgttgcgtt cgtgtcgactt gggccatcggtt ttccaaactctt gttggccgtt
30 7801 caccgttgcgtt cccgggttgcgtt cggccatcggtt ccacacccg gggcgttgcgtt
7861 gggccgttgcgtt cccgggttgcgtt cggccatcggtt ccacacccg gggcgttgcgtt
35

7921 cggctgcgg cggcacatgg cggccatgt ctcaccggcg aatacctgtg acggcgcgaa
7981 atgatcacac ctgcgttct tcgccccgtc agcacatggc gccgctgggc ggcgtggtgc
8041 cggcgattaa atacgcccgc acgtactcgtaatgcagct gtcgccccgg aataccaccg
5 8101 tgtgcgtgggt tccgtcgaag gtcagcaacg aaccgcgaag ctggttcgcc aggtcgaccc
8161 cggccgttgcgttgc cggcgatggc gggatcggatac caccaccgtc ggcactaggc
8221 cggccgcgcga gacggcatgg ggctgacttg tgggtggcac cggccagaac gcgcagggtc
8281 ccagcgccgc atacccgggt aacttcccg agctcatgaa cggtgcatc tcccgccgc
8341 ggcggcttc gtcgtgacc ttgcgtcgat cggtaaccgg gggctgatcg acgcaattga
10 8401 tcgcccacccg cggcgtcaccg gaattgtgt agcggccgtg cgagtcccgaa cgcattaca
8461 tgtcgccag agccagcagg gtgtctccgc gattgtcgac cagctccgac agcccgctgg
8521 tcaagtgttgc acagatttc ggtgactaca ggcgcataat ggtgcccacg atggcgctgc
8581 tataactcag cccgcgcga tccgtcgcc ggcgcggcct gctgatccgc gggttgtccg
8641 ggtcgaccaa cggatcgacc aggctgtggt agacctcgac ggcttggcc gggtcggcgc
15 8701 ccagcgccgc gcccgcgttgc ttggcgact cggccgcata gttgtgaac ggcgttgcga
8761 agcccttggc ctggcgccagc tccgcctcgat tggatcgcc attggggctcg acggcaccgt
8821 cgagaatcat tgcccgacc cgtcgccgaa attcctcgcc atacgcccgg cccatccgg
8881 tgccgtacga gtagcccgagg taggtcagct tgcgtcgcc caacgcgcg cgaatggcat
8941 ccaggccctt ggcgacgttgc accgtcccgat gggccagaa aagttcttgc cccatcttgc
20 9001 ccacacagcg accgcacatgt cgttgcgtat gtcgtccaca ccctccggc
9061 tgttagtcaac ctgcggctcg gcccgcagcc ggtcggtgtc ggcacatggat ttgcaccaga
9121 tcgcccggccg ggacgacgccc accccgcggg ggtcgaaaccc aaccaggatcg aacccttgc
9181 gcacccgcgttgc cggcaatgtc tggaaagacgc ccaaggccgc ctcgatccgc gattcgccgg
9241 gtccaccggg atttatgacc agcgaaccga tcttgcgtcc cgtcgccgaa aagcgaatca
9301 ggcgcaggcgc cggccacgtca ccatcgccggc ggtcgtagtc gaccggatc ggcgcgttgc
25 9361 cgcataacgc gcccgcgggg atctttactt gcggtttga cgcacggcgc ggtgtccact
9421 ccaccggcttgc cggccacgttgc ggcgtccgc tacgagcggc tcccccgcacc acgcggatgc
9481 agcccaacaa aaccaacgc accggccgcgaa ggcgcggccca gatcaacacgc atgcgcgcga
9541 tcttgcgtcgcc ggcgcacgc ctcatggccaa atcgatcgcc agacgcaccc cggatcttgc
30 9601 ggcgcggcc accgtcgccgactaaccgg cccgtccacg cagtcgtccatcgccatcgatgc
9661 ggcgcacgc cggccatccc ccatacgatcc ggtacacatcgcc gccggcaaga caccgcaccc
9721 tcgaccggat cccgcacggg cgcgtccgc tccggccgtgc acaactcgatcgcc catcgatcg
9781 ggcgcgttgc cccgtccacg cccgtccacg gtcgtccgc tccggccgtgc acaactcgatcg
9841 cgcgtccgc cccgcacggg cccgtccacg gtcgtccgc tccggccgtgc acaactcgatcg
9901 tgtggccgcac gatctccgag agtgcggccgc gccgcggccgc ggcgcgttgc aacccgcaccc
35 9961 gcaagtgcacg atcgatgtga ggtcaaggat tcagcgcact gtcgtccgc aacccgcaccc

10021 accgcggcct cgccgtgatc tggagtcaga acgcgtcacg cagccggta aaggcgtaac
10081 ccatgtctcg acaaacatgc atggcgtag tggacgtttc cagacacagc aactggcg
10141 caggccactg agccgctgca tgcgcgatgg tatgcgcgtt gggccccgg ggcgcgtctga
5 10201 ggggaagaag tggcagactg tcagggtccg acgaacccgg ggaccctaac gggccacgag
10261 gatcgaccgg accaccatta gggacagtga tgcgtcgacta gactatctat ggggccaata
10321 ccccccggagg ctccggcccg cggaccaaga tccgcacca ccacctacag agatggaagg
10381 ccgacggcca caagtggcc atgctgacgg cttacgacta ttcgacggcc cggatctcg
10441 acgaggccgg catcccggtg ctgcgtcg tgcattcgcc ggccaaacgtc gtgtacggct
10 10501 acgacaccac cgtgcgatc tccatcgacg agctgatccc gctggccgtt ggcgtgggt
10561 ggggtgcccc gcacgcactg gtcgtcgccg acctgcccgtt cggcagctac gaggccgggc
10621 ccaccggccgc gttggccgccc gccacccgggt tcctcaagga cggccggcga catgcggta
10681 agctcgaggg cggtgagcgg gtggccgagc aaatgccttgc tgcgtcgcc gccccatcc
10741 cggtgatggc acacatggc ttaccccgca aaagcgtcaa caccctggc ggcttccggg
15 10801 tgcaggccg cggcgacgccc gccgaacaaa ccatgcggca cgcgtcgcc gtcggccgaag
10861 ccggagcggt tgccgtcgatgg tggagatgg tggccgcga gttggccacc cagatcaccg
10921 gcaagttac cattccgacg gtcggatcg ggcgtggcc caactgcgac ggccagggtcc
10981 tggatggca ggacatggcc gggttcagcgc ggcggccaaacgc cggccgttc gtcaacgg
11041 atgcccgtgt cggtggtgaa ctacggcgatcg tgcgtcgatggc atacggccaa gaggtggccg
20 11101 cgggggtatt ccccgctgac gaacacagtt tgcgtcgatggc cgcgtcgcc cggatcgg
11161 ggcattgggg tggccgcctg gatgcgtcg acgcccggatt gccggccgg acgcggccagc
11221 gggacccatc ggcgtcgatcg tgcgtcgatggc agccgggggt gagcccagac attcgatgt
11281 cccaaacacca tccggccacag cccaaatggat gtcggactct atgcgtcgatgg
11341 aaccaccacc gggcgacgc atcatgaccg gaggcgaaga tgccagtaga ggcggccaga
25 11401 ccagcgccgc atcggaggt cgagcgcaag tgcgtcgatggc tgcgtcgac ggtgtcgccg
11461 tgcgtcgatgg gcatgcggcc ggtgggtcg tgcgtcgatggc cggccggatcc gcaacgtcg
11521 cgggtgtact tgcgtcgatggc tgcgtcgac ctggccggca accagatcac ctggccgg
11581 cgcacccggcg ggcggccacgc cggctggcat tgcgtcgatggc cggccggacc cggacacgg
11641 accgagatgc gggccggatcc gtcggatcc ggcgtcgatggc tgcgtcgatggc gttgtggat
30 11701 tgggtgtcgatggc cggatcggatcc cggccggatcc gtcggatcc gtcgtcgatggc
11761 cggccaaacggcc agatcgtcgatggc cggccggatcc gtcggatcc gtcgtcgatggc
11821 gacgtcaccgc catggcgatcc cggccggatcc cggccggatcc gtcgtcgatggc
11881 ggcggccacgc agtggcgatggc atggcgatggc gtcgtcgatggc cggccggatcc
11941 accaagctac tggacccggatcc agccaaacggatcc tgcgtcgatggc cggccggatcc
35 12001 cacggctcca aactggcgatcc cggccggatcc gtcgtcgatggc cggccggatcc
12061 cggccggatcc cggccggatcc agtacaccgc gtcgtcgatggc cggccggatcc

12121 ctgtgggatc gggccgtgcg ggccgacgcc tatgacgccc tgcaccagat gcgagtgacg
 12181 acccgcaaga tccgcagctt gctgacggat tcccaggagt cgtttggct gaaggaaagt
 12241 gcggtggtca tcgatgaact gcgtgagctg gccgatgtcc tggcgtagc cccggacgcc
 5 12301 gaggtaactcg gtgaccgcta ccagcgcgaa ctggacgcgc tggcccgga gctggtaacgc
 12361 ggccgggtgc gcgagcgcct ggttagacggg gcgcggcgac gataccagac cgggctgcgg
 12421 cgatcactga tcgcattgcg gtcgcagcgg tacttccgtc tgctcgacgc tctagacgcg
 12481 ctgtgtccg aacgcgccta tgccacttct ggggaggaat cggcaccgg aaccatcgat
 12541 gccgcctacc ggcgagtcg caaagccgca aaagccgcaa agaccgcgg cgaccaggcg
 10 12601 ggcgaccacc accgcgacga ggcattgcac ctgatccga agcgcgcgaa gcgattacgc
 12661 tacaccgcgg cggtacttgg ggcggacaat gtgtcacaag aagccaaggt catccagacg
 12721 ttgcttagcgc atcatcaaga cagcgtggc agccggaaac atctgatcca gcagggcata
 12781 gccgcgaaca ccgcggcgaa ggacacccttc acctacggc tgctctacca acaggaagcc
 12841 gacttggccg agcgtgcgcg ggagcagctt gaagccgcgc tgccaaact cgacaaggcg
 15 12901 gtccgcataaag cacgggattt agccgcgcag gggcgacga gttggctgt aagccggatt
 12961 ctgtccgcg ccgcacacgc caagctaacg gccgcacggc ggcgaccatc catctggaca
 13021 caccgttacc ggggcctcg agcggcctac ccgcaggctc gggcgagcaa ccctcaagcg
 13081 cctgcgcggc cgcaacttcg gtgcggcctt ctggccctg ctgcgggtgg gggttgccta
 13141 gccaccccggt tcacccggaa tgctggcgc ctcttaccgc accgttacac cttgcacc
 20 13201 acgaggatgg cggtctgttt tctgtggcac ttcccgca gtcacctcg attgcgtta
 13261 gcaatcaccc tgctctgtga agtccggact ttctcgact cgacgctgaa cctcgtgaat
 13321 ccacacaagc cctacgcgag ccgcggccgc ccagccaact catccgcgcac gaccacgcta
 13381 ccccgctggg cggtgtcgcg gccagtgtga ccgtggacg acacggctag tcggacagcc
 13441 gatccggcgg gcagtcccta tgcgtggactg gtgacacggt gggacaaacg cgtcgactcc
 25 13501 ggcgacttggg acgcacatgc tgccgagggtc agcgagtaacg gtggcgcact gctacctcg
 13561 ctgatcaccc ccggcgaggc cgccggctcg cgcaagctgt acgcccacga cggccctgttt
 13621 cgctcgacgg tcgatatggc atccaagcgg tacggcgccg ggcagtatcg atatttccat
 13681 gccccatc ccgagtgtac gacggcgcgt gtatccaaa ctgctgccga
 13741 tagcgcgcata ctgggggcc aaactgggcc gggaggcgcc ctggccagac agccttgatg
 30 13801 actgttgtgc gagctgtcat gccgcggcc aaacccgatc cacagcgtcg atgtgaagt
 13861 acggcaccaa cgactggaaac gcccatacc accgtatcta cggcgagttt gtgttccgc
 13921 tgcaggtggt gatcaaccc tgcgtggactg aacccgacta caccggcgcc gagttccgc
 13981 ttgtcgaaca gcggcctcgcc gccaatccc ggggtaccgc aatgcaactt ccgcaggac
 14041 atggttatgt gttcacgacc cgtgatcgcc cgggtgcggac tagccgtggc tggcggcat
 14101 ctccagtgcgcc ccatgggttc tgcactattc gttccggcga acgctatgcc atggggctga
 35 14161 tcttcacga cgcagcctga tgcacgcca tctatagata gcctgtctga ttccaccaatc

14221 gcaccgacga tgccccatcg gcgtagaact cggcgatgtc cagcgatgcc agatcaagat
 14281 gcaaccgata taggacgccc gacccggcat ccaacgcac cgcgaacaac atttgatcg
 14341 gcgtgacatg tgacaccacc agcaccgtcg cgccitcgta gccaacgatg atccgatcac
 5 14401 gtccccggc aacccggccgc agcacgtcg cgaagcttc cccacccggg ggcgtgtac
 14461 tgggtcctg cagccagcga cgggtcagct cgggatcgcg ttctgcggcc tccgcgaacg
 14521 tcagccccctc ccaggcgcgg aagtccgtct cgaccaggc gtcatcgacg accacgtcca
 14581 gggccaggc tctggccggc gtcacccggg tgcgttaagc cccgtgtac ggcgaggaga
 14641 ccaccgcgc gatcccgccg cgccgcgc aataccggc cgccgcacca acctggcgcc
 10 14701 accccaccc gttcaacccc gggttgcgc gccccgaata gccgcgttgc tccgacagct
 14761 ccgtctgccc gtggcgcaac aaaagtatgc gggtgggtgt accgcggcgc cccgtccacg
 14821 cgggagatgt cggtgactcg gtcgcaacga ttggcagg atccgcatcc gccgcagccg
 14881 attgcggc ggcgtccatc cgcgtattgg ccaacccggc tgcatacg ttcgggac
 14941 gccgaaccca ctcgtatgt atcctgcgaa actgggacgc caacgcctga gcctggacat
 15 15001 agagcttcag cagatccggg tgcgttgcct tccaccggcc ggacatctgc tccaccacca
 15061 gcttggatgc catcagcacc gcccgcctcg tggcacctag ttacccggc tgcgtccaaac
 15121 cggctatcag gcccggat tccggcgttgc tgcgttgc cccgcgcgc gcctgcgttgc
 15181 actccggcag cacgggtggag tgcgttgc tccacccac cccgcgtat cccggggc
 15241 cgggattgcc cccgcgcgc cccgcgttgc cgcgttgc cccgcgcgc
 20 15301 gagccgcaac aagatcgtc cgcattccgg gcagcgcacc acttcatttc cggcgccgc
 15361 cggatctgg cccgcgttgc cccgcgcgc cccgcgttgc cccgcgcgc
 15421 ttcaaccgc cccgcgcgc cccgcgttgc cccgcgcgc
 15481 ctcggatca agtgcgttgc tccacccac cccgcgtat cccggggc
 15541 gtcgatttcg gcaagtgcct cgtccaaac cccgcgttgc cccgcgcgc
 25 15601 cgcgttgc cccgcgtat cccgcgttgc cccgcgttgc
 15661 cacccgcgcgc cccgcgttgc cccgcgttgc cccgcgttgc
 15721 ctcggatca gcaatttgc tccacccac cccgcgtat cccggggc
 15781 gtcgatccgc tccacccac cccgcgttgc cccgcgtat cccggggc
 15841 gtcgatccgc tccacccac cccgcgttgc cccgcgtat cccggggc
 30 15901 caccgttgc tccacccac cccgcgttgc cccgcgtat cccggggc
 15961 ctcggatcc gcaatttgc tccacccac cccgcgttgc cccgcgtat cccggggc
 16021 ctcggatcc tccacccac cccgcgttgc cccgcgtat cccggggc
 16081 cgcaccggca gcaacactt cccgcgttgc cccgcgtat cccggggc
 16141 gggatttcg tccacccac cccgcgttgc cccgcgtat cccggggc
 16201 tcgttgcgttgc gatgtatgc cccgcgttgc cccgcgtat cccggggc
 35 16261 acggatcc gcaacactt cccgcgttgc cccgcgtat cccggggc

16321 tcgggatccc cggcgccgac cacaccggtc gcagtcggcg gcaacgcggc ctccagacgg
16381 gcaacaaagg tgcgcagcgg ttgggtttt ggcagtctgc caatccggcc taacccgctg
16441 ccgaccggcg gtggtaccag cgcaagatg tcgaatgcgg gtcctcgta agggtgcgcg
5 16501 ggcgcacatcg ccgccaacac ctcggcgcgc gtcgtgcgg gtgcgacgac ctcgacccgg
16561 tcctcgccca cccgttcgac ggtaccgacg ctgcctatgg cgggcacgc cccgtcg
16621 gccaggaact gcccggtacc cgacactc cagctgcagt gcgagtagtc gccgatatgg
16681 ccggcaccgg cctcaaagac cgtgtccgc accgcctctg agtctcgcg cggcacatag
16741 atgacccact tgcgagatc ggccgctccg ggcacccggg cgagaacggc gtcgacggc
10 16801 agaccaacag cgtgtgcgcg cgcgtcgac acacccggcg acgcgcgatc ggcgttgg
16861 tgcgcggtaa acaacgagcg accggtccgg atcaggcggt gcaccagcac acccttggc
16921 gtgtggccg cgaccgtatc gaccccacgc agtaacaacg ggtggtcac caatagcagt
16981 ccggccgtgg gaacctggc caccaccgc ggcgtcggt ccaccgcaac ggtcaccgaa
17041 tccaccacgt cgtcggttc gcccacacc agacccaccc aatcccacga ctggcaagc
15 17101 cgcggcgggt aggccgtgc cgcacgtcg atgacatcg ccagccgcac actcatcg
17161 gtcctccacg ctttgcac tggcgatcg cgcaccagg cacggccac tccggcgca
17221 cgcgcggcc caggtaccgc ggcgtccaggc cgacgaagg gtcaccgcgg cgcaccgcaa
17281 tcccttgcgtc cgtcaatag ttgcataatc cgtcagcatc ggcgtatgg aacagtacga
17341 aaggggccgc accatcgacc acctcgaccc caccgcgtc cgtccggcc accatctcg
20 17401 cgcgcagcgc cgtcaaccgc accgcacatgg ctgcggcagc ggcgcaccgc cggggggcgc
17461 agcaagcagc gatggccgtc agttgcataatg ttcccaacgg ccagtgcgt cgctgcacgg
17521 tcaaccgagc cagcacgtct ggcgagccga ggcgttagcc cacccgaat cggccagcg
17581 accacgtttt cgtcaagcta cggagcacca gcacatcggtt cagcgagtca tcggccaacg
17641 attgcggctc gccgggaacc caatcagcga acgcctcgatc gaccaccagg atgcgtcccg
25 17701 gcccggtaa ctcgagcage tgcgtcgatc ggtgcagcac cgagggtggg tttgtcgat
17761 tacccacgc gacaagggtcg ggcgtcgatc gcacgtcgatc ggtgtccagg acgaacggcg
17821 gctttaggac aacatgggtc ggcgtcgatc cggcagcgatc caaggctatg cccggctcg
17881 tgaacgcggg cacgacgatt gtcggccga cggacttag gttgtcgatc aatgcgaatc
17941 ctcggccgc cccgacgac gggaggactt cgtcaccggg tctgccaatgc cgttcagcga
30 18001 cgcgtcttg cgcgggtgc acatcgatgg tgcgtcgatc gcccggccagg tccggcagca
18061 ggcgcggcag ctgcggacc aaccattccg gggggccggc atggcggacg ttgcggcga
18121 agtccagcac gcccggccgc acatcgatc caccgtggta ggcgcggccgc gcaaggccgg
18181 tagtgcgtatc actcgccaca ggcgtcaaaaca gttgtggcc ggtgtcgatc ccaagaatcc
18241 agagcaccgc cgcgtcgatc tctacgcggc gacaaccgcg acatcacagg cagctaaacag
35 18301 ggcgtcgatc gtcggatcg tcaggccaaatc cagctgtgcgc tggcgtatgc gcaacacgg
18361 gaatggatgt cgtatggatc ccggaaatc tgcgtgcgc agtgcgtgcg tggcgtatgc
18361 gaatggatgt cgtatggatc ccggaaatc tgcgtgcgc agtgcgtgcg tggcgtatgc

18421 acagcggcga cgtgccgcag cggcgattc gatcgggcac gtaagaagcc gatggctcg
 18481 gccccggag ctgccgagg cggtagtta tcgcgatctc ccaggactg gccccgaca
 18541 agagaatgct gttcggacg tcctgaacaa tcgcccgtt ttcgttacg gcatccgcag
 5 18601 ccaaactgtgg gtgtcgatga ggtacgcgtt caccggtaa agcgttcgag cacgtcgct
 18661 gacaacggag cgtccaaatc gtcgggcacg cggtagacgc catggtaat gcctaaccgc
 18721 cgagtctcat gaggatgcag cggcacaagg tttgctaccg gtcggccgcg gccccgcaatc
 18781 tcaacctctg cccggcttag acgagccgca gcagctcgaa caggcggttc ttgcctcg
 18841 gaacgcccac cccgttcgca ggcggccaga cttcgcgtc gaccacactgc tcaccaaact
 10 18901 tcgcgatcat cgcctgatac cacagcgcca acgggtacg gtttgtccaa cgcgttcgtc
 18961 aacgacaatg ggatcgacgacgacgacc gcgagcggga ccaattgccc gcctccctcca
 19021 cgcggccgcg cacggcgccgc atcgtcgccg ggtgaatcgc cgcagctggatctcgat
 19081 ctggacggca cgcgtgaccga ctggcgccgc ggaatcgat ccagcttcg acacggcgtc
 19141 aaccacatcg gtggcccaatcccgaaaggc gacctggcca ctcacatcg cggccggccc
 15 19201 atgcgtgaga cgcgtcgccgc catggggcgc ggcgaatcgc cggaggaggg gatgttagcc
 19261 taccggcccg actacagcgc cgcgggttgg gcgatgaaca gcttgtcga cggatcgccc
 19321 cgcgtcgatgg cgcgtcgccgcg caccggccgtt gtcggctgg cgcgtccac ctccaaggca
 19381 gagccgaccg cacggcgaat cctgcgcac ttcggaattt agcagcactt cgaggatcatc
 19441 gccccggcga gacccgttgg ctcgcgaggc agcaaggatcg acgtgtggc ccacggcgtc
 20 19501 ggcgcgtgc ggcgcgtacc cggcggttgg gtcgtggcgc ggcgcgttgc ccacgcgtc
 19561 gacggggccgg cgcgcacgg catcgacacg gtcgtggcgc gtcggggcta cggccgcgc
 19621 gactttatcg acaagacatc caccacgtc gtcgcgtat ccgcacatcg tgacgagatcg
 19681 agggaggcgc taggtgtctg atccgtcga cgtcacattt gttgtacgg gcaacatctg
 19741 cgggtcgcca atggccgaga agatgtcgc ccaacatcg ccgcacatcg gctgggtga
 25 19801 cgcgtcgca gtgaccatcg cgggcaccgg gaaatggcat gtggcgttgc ggcgcgttgc
 19861 gccccggcc ggggtgttc gggccacgg ctaccctacc gaccaccggg cgcacaatgt
 19921 cggcaccgaa caccggcgccg cagacatgtt gtcggccat gaccgcaacc acgctcggt
 19981 gttcggcag ctggcgatcg aagccgcgg ggtacggatg ctgcgttgc tcgacccacg
 20041 ctggcggaccatcgatcgatgtcgatcgatccctactat ggcgttgcact ccgcgttcga
 20101 ggaggatcttc ggcgttgc aatccgcctt gcccggcttgc caccgttgc gtcgttgc
 20161 ttcgtcgccgcg aacggaccga gttgttgc gtcgttgc gtcgttgc gcccggcttgc
 20221 ctggcgatgg ccctggcgatgttgc gtcgttgc acctacatgt gtcgttgc gtcgttgc
 20281 tggcgttgc gcaagaatgc caaaacgtca cgagagaacc agcagatcgatcgatcg
 20341 gacacccgcg cggatccgcgatgttgc gaaatccctt ctaccacatcg aggatcgatcg
 20401 ggcgttgc gccgggttgc ggcgttgc gtcgttgc gtcgttgc gtcgttgc
 30 20461 cgactcgatcgatcgatcgatcgatcgatcgatcgatcgatcgatcgatcgatcgatcg
 35

20521 ggccggaccaa ccgtcctggc cgaccgtgga tacgtgcggc cccaggtggg ctgcacgta
 20581 ccaccgatcc cccgcctgccc ggtgcagacg gtgaccatca cggcgggct gcggtactcc
 20641 gaaccgagcg tggggggcaa agaccattc gtcagagacg gcttccagca ggtgtattcg
 5 20701 atcaataccg gacaggtcgc cgcgctgacc ggagtccagc tggctgggtc ctatctgcag
 20761 ttgatcgaag accaaccggg cgggctcgcc gtcgctggcg ttccgcacatc agatccggg
 20821 ccgttcctgt cctatggcat ccaatggatc tcgttcggca ttctggcacc gatcggttg
 20881 ggttatttcg cttacgcgca gatccggcg cggccgggg aaaaaggcggtt gtcgcccacca
 20941 ccggacaagc caatgacggt cgagcagaaa ctgcgtgacc gctacggccg cggccggtaa
 10 21001 accaacatca cggccaaatac cgcaaaaaaa gcctggacca cccggacag caccacggcg
 21061 cggcgagat cggccaccctt gggcgaccgg cggtgcacca aggtggcccg gatctgcaac
 21121 tcatgggtt accgggtggg cccaccggc cgcacgtcaa gggccggc aaacgcccgc
 21181 tcgacgacac cggcggtggg gtcggatgg cggggccgt cggccgcacca ggccgttacc
 21241 gcaccgcggg gcgacccacc gaccaccggc ggcagatca ccaccagcac cggcgtcgcc
 15 21301 cgtgcgcacaa catagttggc ccagtcatcc aatcgctg cagccaaacc gaatcgaga
 21361 taacgcggcg agcggttagcc gatcatcgag tccagggtgt tgatggcacg atatccacg
 21421 accgcaggca cggcgtcgaa agccgcaccc acgcaggcgc ccacctggc gtcggcggtg
 21481 tttccgcacca cggactccag cggcgcacgc gtcaggcccg gggcccccag ctggccggg
 21541 tcacgcgcgc acagcgacgg cagcagccgt cggccgcct cgcacatcg cgcgtccaaac
 20 21601 aggtccgata tctggcgcc ggtgcgcgc agcgaagttt cggccagcgc tgccaggtg
 21661 gcccgtcgccgg tgccgcaccc gggccaggac ctggccggta gcccgtgcag tgccgcgc
 21721 agcaagccca cccgcgcgcac cagcaggccg acgtgtaccg caccggcgcac cggccgtca
 21781 cggtaggtga tctgctccag ctggcgcc gcccgcacca acaggccac cggatgaccc
 21841 cgtttgggtt cggcgaacac gacgtcgagc aggccgcgc tca gacgcaccc gacggccctg
 21901 gtcgtccagg tggatgcaaa cactccggca gctgcgcaca cgtggctac gtcagctat
 25 21961 ttatgaccc atacggcagc tatccacgtt gaagcgccca gtcacccggg ttggccaccc
 22021 gttgaaccccg gggcaatgt tggccggc acgcaatgtc atcatgcagc tggcgtgc
 22081 ggggtcggtt tatggcgtgc tggaaagccc ggtggacacg ggcaacgtct acaagcatcc
 22141 gttcaagcg gcccggacca cggcaccta cttggcggtt ggcacccatcg ggacggaaatc
 22201 cgaccgagcg ctgatccggg gtcggcgtt gtcgcgcac cggcagggtt gtcgcacggc
 22261 ctgcaggccca gtgtccata acgccttcga cccgaagttt cagctgtggg tggccggcg
 22321 tctgtaccgc tacttcgtt accagcacga gtttctgtac ggccactcg aagatgcac
 22381 cggccgcacgc gtcaccaag acgcacccac gtttagggacc acgctgcagg tgccggagg
 22441 gatgtggcccg cggacccggg tcgcgttcga cgagttactgg aagcgctcg tttatgggt
 22501 gcaagatcgac ggcgcgggtgc gcgagcatct tcgcgggtt ggcacccatcg ctttctcc
 30 22561 gtggccgtt cggcgggtt cggccgtt caaccgttt ggcacccatcg gatttcttggc
 35

24721 tccgacgatc cccatcgca gcagcgcat gcagccgagc gccagggcgc ctgcctgg
24781 gagctccccc ccgtgcctgc gaggcacggc gcgcacatcg atgacacgg catgtgagg
24841 tacctggtcg cagcgacacc gcgcgtggcg tgggtgtcg cgcatccgca gaaccgagcg
24901 gagtgccgct atccgcccgc gacgcgggtg cggcacgata gggggacgac catctaaaca
24961 gcacgcacgc ggaagccgc cacctacagg agtagtgcgt tgaccaccga ttgcggccgc
25021 cacgatctgg cccaaaactc aaacagcgca agcgaacccg accgagttcg ggtgatccgc
25081 gagggtgtgg cgtcgatatt gcccgcacatt gatcccgagg agacctcgga gtggctggag
25141 tccttgaca cgctgctgca acgtgcggc ccgtgcggg cccgctacct gatgtgcgg
25201 ctgctagagc gggccggcga gcagcgggtg gccatcccg cattgacgac taccgactat
25261 gtcaacacca tcccgaccga gctggagccg tggttccccg ggcacgaaga cgtcaacgt
25321 cgttatcgag cgtggatcgat atgaaatgcg gccatcatgg tgcaccgtgc gcaacgaccg
25381 ggtgtggcg tgggtggcca tatctcgacc tacgcgtcg cccggcgct ctatgaggc
25441 ggtttcaacc acttctccg cggcaagtcg caccggcg gcccgcata ggtgttcatc
25501 cagggccacg ctccccggg aatctacgcg cgcgccttcc tcgaaggcg gttgaccgccc
25561 gagcaactcg acggattccg ccaggaacac agccatgtcg gccggcggtt gccgtctat
25621 ccgcacccgc ggctcatgcc cgactctgg gaattccccca ccgtgtcgat gggttggc
25681 ccgctcaacg ccatctacca ggcacgggtc aaccactatc tgcattgaccg cggtatcaaa
25741 gacaccccg atcaacacgt gtgggtttt ttgggcgacg gcgagatgga cgaacccgag
25801 agccgtggc tggccacgt cggcgcgtg gaaggctgg acaacttgc ctctgtatc
25861 aactgcaatc tgcagcgact cgacggcccg gtgcgcggca acggcaagat catccaggag
25921 ctggagtctgt tctccgcgg tgccggctgg aacgtcatca aggtgggtg gggccgcgaa
25981 tggatgccc tgctgcacgc cgaccgcgac ggtgcgtgg tgaatttaat gaatacaaca
26041 cccgtatggcg attaccagac ctataaggcc aacgcacggcg gctacgtcg tgaccactc
26101 ttggccgcg acccacgcac caaggcgctg gtggagaaca tgagcgacca ggatatctgg
26161 aacctcaaac gggcgccca cgattaccgc aagggttacg ccgcctaccg cgccggcg
26221 gaccacaagg gacagccgac ggtgtccctg gccaagacca tcaaaggcta cgcgtggc
26281 aagcattcg aaggacgaa tgccacccac cagatgaaaa aactgaccc tgaagaccc
26341 aaggagtcc gtgacacgcgac gcggattccg gtcagcgacg cccagcttga agagaatccg
26401 tacctgcgc cctactacca ccccgccctc aacgcggcccg agattcgta catgtcgac
26461 cggcgccggg ccctcgaaaa ctttgttccc gagcgcagga ccaagtccaa agcgctgacc
26521 ctggccggcgc ggcacatcta cgcgcgcgtg aaaaaggctt ctgggcacca ggaggtggc
26581 accaccatgg cgacggcgac cagctaaaa gaagtgtgc ggcacaaagca gatcgccgg
26641 cggatagtcc cgatcattcc cgacgaggcc cgcacccgt ggtggactc ctggttcccc
26701 tcgtctaaaga tctataaccg caatggccag ctgtataaccg cgggtgacgc cgacccatg
26761 ctggccctaca aggagagcgaa agtccggcag atcctgcacg agggcatcaa cgaagccggg

26821 tcgggtggct cgttcatcgc ggccggcacc tcgtatgcga cgcacaacga accgatgatc
 26881 cccatitaca tcttctactc gatgttcggc ttccagcgca cccgcgatag cttctggcc
 26941 gcggccgacc agatggctcg agggttcgtc ctggggcca cgcggggcg caccaccctg
 5 27001 accggtgagg gcctgcaaca cgccgacggt cactcggtgc tgctggccgc caccacccg
 27061 gcgggtggt cctacgaccc ggccttcgccc tacgaaatcg cctacatcggt ggaaagcgga
 27121 ctggccagga tgtgcgggga gaacccggag aacatcttct tctacatcac cgtctacaac
 27181 gagccgtacg tgccggccgc ggagccggag aacttcgatc ccgagggcggt gctgcggggt
 27241 atctaccgct atcacgcggc caccgagcaa cgcaccaaca agggcgagat cctggctcc
 10 27301 ggggttagcga tgccggccgc gctgcgggca gcacagatgc tggccgcccga gtggatgtc
 27361 gcccggacg tgggtcggt gaccagtgg ggcgagctaa accgcgacgg ggtggccatc
 27421 gagaccgaga agtccgcca ccccgatcgcc ccggccggcg tggccctacgt gacgagagcg
 27481 ctggagaatg ctggggccgc ggtgatcgcg gtgtcgact ggtgcgcgc ggtcccgag
 27541 cagatccgac cttgggtgcc gggcacatac ctcacgttgg gcacccgacgg gttcggctt
 15 27601 tccgacactc ggccggccgc tcggcgtac tttcaacaccc acgcccatac ccagggtggc
 27661 gcgggtttgg aggccgtggc gggcgacggc gagatcgacc catcggtccc ggtcgccggcc
 27721 gcccggcagt accggatcga cgacgtggcg gctgcgcggc agcagaccac ggatcccggt
 27781 cccggggcc aacccggcg agccgaccgc cttggccga atctccaga aatctggcg
 27841 agcttttagg agtgaacgc aatcagttgg ctccagttgc ccggccgagg tggccgtcg
 20 27901 aactgctgga cactgtcccc gattcgctgc tgccgggtt gaagcgtac tggccggcc
 27961 tggccaccga ggcagttcg gccatgcaag aacgggtggc gttttcgcc gacctagaag
 28021 cgtcccgacg cgcacgcgtg ggcgtgggg tgcagacggc cgtggtaac ttcgtcaat
 28081 ggtgcacga cccgcacagt gacgtcggt ataccgcga ggcattcgag ctgggtcccc
 28141 aggtctgcgac ggcacggatc ggcgtgcgc agaccgtgga catggtcgg gtcaccatgg
 25 28201 agtcttcga agaagtcgtg cccctgcgtg cccgttccga agacgttgg accgcctca
 28261 cgggtggcat ttgaaatac agccgcgacc tggcattcac cggccacacg gcctacgcgg
 28321 atgcggccga ggcacgaggc acctggaca gccggatgga ggcacgcgtg gtggacgcgg
 28381 tggtaacgcg cgcacccgtt cccgagctgc tggccgggc ggcgcgtg aattggaca
 28441 ccaccgcgc ggcgaccgtt ctggggaa ctccggccgc cggtaaat ggctccaaca
 30 28501 ggcacggcga cagcgagcgg gccagccagg atgtccgcga caccgcggct cggccacggcc
 28561 ggcgtgcgt gaccgcgtg cacggcacct ggctgggtggc gatgtctcc ggcacgtgt
 28621 cggccacgcg gaagttccctc aaagacgtc tggcagcatt cggccacgc cccgtggta
 28681 tggcccccac ggcgcacatc ctgaccgcgg cgcacccgtg cgtacgcgag ggcacgtcc
 28741 ggtgaacgc cgtcgccggc tggcgccggag cgcgcggcc cgtgtggct agggacttt
 35 28801 tggccgaacgc cgcctgtatc ggcgacgcct cggcgatgtt ggcctgcgtt accgcacgtga
 28861 tggcccccctt agccgatgcc ggcacgcgtc tcatcgagac gctacgcgtatctggatt

28921 gtggcggcgc gattgaagct tggccagaa agttgtcgt tcatccaaac acagtgcgg
 28981 accggctcaa gggatcacc gacttcaccc ggcgcgatcc caccagcca cgcgcgc
 29041 atgtcctcg ggtggcggcc accgtggtc aactcaacta tccgacgccg cactgaagca
 5 29101 tcgacagcaa tgccgtgtca tagattccct cgcggcgtcag agggggtcca gcagggccc
 29161 cgaaaagata ccagggcgc cgtggacgg aaagtgtatcc agacaacagg tcgcggacg
 29221 atctaaaaaa catacgttac aggcccgtt tttttttt atacaaaaac ctaagacgag
 29281 gttcataatc ttttacaccg cgcggacgg aaagtgtatcc agacaacagg tcgcggacg
 29341 gcgttgctcg caccggaca ggggtcgcaa accgaggaa ttttgcgcgtt gttgcgc
 10 29401 ctggccggcg cagcggacca gatcgccggc tggcgaaag cgcgtatct agatctgc
 29461 cggctggca ccaccgcctc gaccgaggat atcaccgaca cgcggcgcgc ccagccatt
 29521 atcgtcgccg cgactctgtt gcccaccag gaactggcgc gccgtatcgt gtcgcgc
 29581 aaggacgtca tcgtggccgg ccactccgtc ggcggaaatcg cggctacgc aatcgccgt
 29641 gtgatagccg ccgacgacgc cgtcgccgtc gccggccaccc gggcgccgaa gatggccaa
 15 29701 gcctgcgcca ccgagccgac cggcatgtc tgggtgtcg gggggacga gaccgagg
 29761 ctgagtcgccc tcgaggatcg ctactggtc cggcaaaacc gcaacggcgc cggccagatc
 29821 gtcgtgcgcg gcccgtac cgcgtggag aagctgcgcg aagaccgcg gccaaggcg
 29881 cgggtgcgtg cactgggtt cgcggagcg ttccacaccg agttcatggc gcccgcactt
 29941 gacggcttg cggggccgcg gccaacatc gcaaccggc accccaccgc cacgtgcgt
 20 30001 tccaaaccgcg acgggaagcc ggtgacatcc gggccgcgg cgatggacac cctggctcc
 30061 cagctcaccc aaccggcgcg atgggacatc tgcaccgcga cgatgcgcga acacacatc
 30121 acggcgatcg tggagttccc ccccgccggc acgttgcgcg tgcgcgcaccc acgcgaaactt
 30181 cgggggggttc cggcacgcgc cgtcaagtca cccgcagacc tggacgagct ggcaaaccta
 30241 taaccgcgga ctggccaga acaaccatc accgtcgtt tcgatttgc cacaacatc
 30301 tacgaaggga agcatgtgt gcctgtact caggaagaaa tattggccgg tattggccg
 30361 atcatcgaag aggttaccgg tatcgacccg tccgagatca ccccgagaa gtcgttcgt
 30421 gacgacccgtt acatcgactc gctgtcgatc gtcgtgcgtc cgcgtgcgcg cggggacaag
 30481 tacggcgtca agatccccga cgaggaccc gcccgtctc gtaccgtcgg tgacgttgc
 30541 gcctacatcc agaagctcgaa ggaagaaaac cggggggccg ctcaggcggtt ggcgcgc
 30601 attggatcg agaaccgcg tgccgtgc aacgttcagg cgaggccgtt ggcgcgc
 30661 aagtgtatcc gcttccacc gtaatggcg gtttcccg cgttgcgtt accggccgtca
 30721 cagcgtacgt gtcgtacgt cggacatcg agacgttgcg gatgggttgc ttggccggc
 30781 agagcggcat ccacgcactc gaagacgatcg tgcgtacccaa tgggatctt ggcgtcaaga
 30841 tcggcgtca ctcaggat cggcgtacca gcccgttggg ccgtacgcg atgcgtac
 30901 tgcgtacgt ccagcggatg ggcaagttgc tggcggaca gctatgggag tccgccc
 35 30961 gcccggatgtt cgttccagac cgggtcgccg ttgttgcgg caccggctt ggtggagccg

31021 agaggattgt cgagagctac gacctgatga atgcggccgg ccccccgaag gtgtccccgc
31081 tggccgttca gatgatcatg cccaaacggfg cccggccggat gatcggtctg cagcttgggg
31141 cccgcgcgg ggtgtatgacc cccgggtcgcc cctgttcgtc gggctcgaa gcgatcgccc
31201 acgcgtggcg tcagatcgatg atggggacg cccgacgtcgcc cgtctgcggc ggtgtcgaaag
31261 gaccatcgaa ggccgtcgccc atcgccggct tctccatgtat gccggccatcgatcgcggca
31321 acgacgagcc tgagcgggccc tcccgccgtatcgatc gacaaggatcgac cccgacggc tttgtgttc
31381 gcgaggccgg tgccgtatcgatc catcgatcgatc gggaggagca cccaaagcc cgtggccgcca
31441 agccgtggc cccgttgcgt ggtccggta tcacatcgatc ccccttcat atgggtggcgc
31501 cccgcggccga tgggtgtcgat gccggtaggg cgtatcgactcg ctcgtggag ctggccgggt
31561 tgcgcggc ggacatcgatc acgtcaacg cgcacggc ac ggcacgcct atggcgacg
31621 cccggggaggc caacgcacatc cccgtcgccg gttgtatca ggcgcgggt tacgcgcgca
31681 agtctgcgtt gggccactcg atcgccgcgg tccgtcgct cggatcggtg ctcacgggtc
31741 tgacgctcgatc cccgcggc atccgcggatc cccatcgatc cccatcgatc cccgcggatc
31801 tcgaccatcgatc cccatcgatc cccatcgatc cccatcgatc cccatcgatc cccatcgatc
31861 actcggttcgg gtcggccggc cacaatgtgg cgttgcctt cggggcgatc tgaaggacatcg
31921 catcgccggatc cccggggggc gaggtggggg tcccccgtatcgatc gggggggcg agtccggaccc
31981 atatggaaagg aacgttcgcata agaccaatgatcgatc cggagctgtt taccggaaa gcctttccct
32041 acgtatcgatc cccatcgatc cccatcgatc cccatcgatc cccatcgatc cccatcgatc
32101 ggaagttgtt gtcggaccgc cccatcgatc cccatcgatc cccatcgatc cccatcgatc
32161 agttcgatc cccatcgatc cccatcgatc cccatcgatc cccatcgatc cccatcgatc
32221 cccggatcgatc cccatcgatc cccatcgatc cccatcgatc cccatcgatc cccatcgatc
32281 cccatcgatc cccatcgatc cccatcgatc cccatcgatc cccatcgatc cccatcgatc
32341 cccatcgatc cccatcgatc cccatcgatc cccatcgatc cccatcgatc cccatcgatc
32401 gaatggaaaggc ggtctcgccg cccatcgatc cccatcgatc cccatcgatc cccatcgatc
32461 cccatcgatc cccatcgatc cccatcgatc cccatcgatc cccatcgatc cccatcgatc
32521 cccatcgatc cccatcgatc cccatcgatc cccatcgatc cccatcgatc cccatcgatc
32581 cccatcgatc cccatcgatc cccatcgatc cccatcgatc cccatcgatc cccatcgatc
32641 tgcgcatcgatc gatgtccacc aacaacgacg accccggccg tgcgcatcgatc cccatcgatc
32701 gggaccgcga cccatcgatc cccatcgatc cccatcgatc cccatcgatc cccatcgatc
32761 agcacgccaatcgatc cccatcgatc cccatcgatc cccatcgatc cccatcgatc cccatcgatc
32821 cccatcgatc cccatcgatc cccatcgatc cccatcgatc cccatcgatc cccatcgatc
32881 cccatcgatc cccatcgatc cccatcgatc cccatcgatc cccatcgatc cccatcgatc
32941 cccatcgatc cccatcgatc cccatcgatc cccatcgatc cccatcgatc cccatcgatc
33001 gcaaccgacc ggcgggttac gcccccaatcgatc cccatcgatc cccatcgatc cccatcgatc
33061 ggcggatcgatc cccatcgatc cccatcgatc cccatcgatc cccatcgatc cccatcgatc

33121 tgaatctggtaaacacctcgatcccgagatcgattggacgtgggggggtgaaccgcac
33181 cgggcaattacggatcgatcaataactcgatcgattcgccggacaaacgtggcaaa
33241 tcgccttcggacggtaactaaccccgagatcgatcgacagagacctcgatgacaatca
33301 tggcccccggagcgggtggcagatcgatcgaccggcgatccgctgtcgctgagca
33361 acttcctcgacgacggcagcgtgaaatgtcgacgacgctgaccgtccggagtg
33421 ccgcggcgggaccgtcaacgggtgcacccatcgatcgatcgaccggcaccgtga
33481 tggcggcgcgcatggcgatcgaggggtgcaatcgatcgatcgaccgtcaacgtgcca
33541 tcgaagaccaatcgatccatcgatggcattcgatcgatcgatcgatcgatcgatcg
10 33601 gtgtgcgggcgtgcacgcgttaggcaggatcgatcgatcgatcgatcgatcgatcg
33661 acatcccgcaatcgatcgatcgatcgatcgatcgatcgatcgatcgatcgatcg
33721 cgatcgatcgatcgatcgatcgatcgatcgatcgatcgatcgatcgatcgatcg
33781 tggtgcgcagatcgatcgatcgatcgatcgatcgatcgatcgatcgatcgatcg
33841 accacaagaaatcgatcgatcgatcgatcgatcgatcgatcgatcgatcgatcg
15 33901 gtggcgcgcgtggatcgatcgatcgatcgatcgatcgatcgatcgatcgatcg
33961 agggcggtgcacatcgatcgatcgatcgatcgatcgatcgatcgatcgatcg
34021 tgcgtccgatcgatcgatcgatcgatcgatcgatcgatcgatcgatcgatcg
34081 attggcgccatcgatcgatcgatcgatcgatcgatcgatcgatcgatcgatcg
34141 tggcaacaaatcgatcgatcgatcgatcgatcgatcgatcgatcgatcgatcg
20 34201 cgcttcgtcgatcgatcgatcgatcgatcgatcgatcgatcgatcgatcgatcg
34261 cggctatctcgccgtgtcgatcgatcgatcgatcgatcgatcgatcgatcg
34321 tgctgcacgcgtcgatcgatcgatcgatcgatcgatcgatcgatcgatcg
34381 acggcggtgcgatcgatcgatcgatcgatcgatcgatcgatcgatcgatcg
34441 ggccggacgcgtcgatcgatcgatcgatcgatcgatcgatcgatcgatcg
25 34501 agaagttggccgatcgatcgatcgatcgatcgatcgatcgatcgatcgatcg
34561 agcatgagcgatcgatcgatcgatcgatcgatcgatcgatcgatcgatcg
34621 agatcgacccatcgatcgatcgatcgatcgatcgatcgatcgatcgatcg
34681 ggccggccgcgtcgatcgatcgatcgatcgatcgatcgatcgatcgatcg
34741 acgcgcgaggatcgatcgatcgatcgatcgatcgatcgatcgatcgatcg
34801 tggtaacgcgatcgatcgatcgatcgatcgatcgatcgatcgatcgatcg
34861 tcgagaaggccgatcgatcgatcgatcgatcgatcgatcgatcgatcgatcg
34921 agatcgacgtatcgatcgatcgatcgatcgatcgatcgatcgatcgatcg
34981 cacaagagccatcgatcgatcgatcgatcgatcgatcgatcgatcgatcg
35041 ccgtcgatcgatcgatcgatcgatcgatcgatcgatcgatcgatcgatcg
35101 ctatccatcgatcgatcgatcgatcgatcgatcgatcgatcgatcgatcg
35161 aacgcgtcgatcgatcgatcgatcgatcgatcgatcgatcgatcgatcg

35221 ctgccgtgga ggtcgacga cagtgcgcc gcccgcggc gctggccacc ctgcacgccc
 35281 cactacggtc aatgcgtgc gctcgactg aaattaaaaa cgccgttgc gagecagcgg
 35341 gcccgcgagg catcgccgcg gcgccgaaac tcttaccctt cgccgacgga cgcgcggaaat
 5 35401 cggccatgga gagcgaggct cggctgtca tgatcgacca cggctgccc ttgcccgaac
 35461 ttcaataccca gatacacggc cacgggtgt aaatgtggcg agtcgactc gcctggcccg
 35521 acatgcgtct cgccggcgaa tacgaaagca tcgagtgcca cgcgggaccg gcccggatgc
 35581 tggcgacaa gacacgctgg gccaagctcc aagagctcggt gtggacgatt gtcccgattg
 35641 tcgtcgacga tgtcagacgcg gaacccggcc gcctggccgc cgcacatgcc cgccacctcg
 10 35701 accgcgcgca tatggccggc tgaccgctgg tgagcagacg cagactcgca ctgcggccgg
 35761 cgcaactgcga ctctgcgtct gctcgctc aacggctgag gaactccctt gcccacggcga
 35821 stacgcgtc gcatcccggtt ggcaccagac cgcacccgggtt ccggccgtcg aggatatcg
 35881 ccacatcccg cgcccccctca tgggtcaccg cgtattcgaa ctccggccgg gtcacgtcga
 35941 tgcgtcgcc gacccggctcg gtggccgct cacatgtggc ggcggcagcg acgttggccg
 15 36001 ctcggccccc gtacccgcacc accagcgact cggcaatcc ggcggcccgat ccggggggccg
 36061 gcccagggtt cgccgggtcgcc cgcgtcgatcg gcagggttgcg agtgcggcac ttgcggctc
 36121 gcagggttgcg cagcgtgtatcg ggcgtcgatcc gcacatccctc tgccatgttag cggattcccg
 36181 tcagcttgcc gccgaccaca ctgatcacgc cgcacggcgat ttcaaaaaca gctgtggtcac
 36241 gcgaaacgtc ggcgggtcgcc ccctggacac cgcacccggcc ggtgtcgatt agcggccgca
 20 36301 atcccgata ggcaccgtatcg acatccctgg tgccgaccgc cgtcccaat gcggtttca
 36361 ccgtatcccg caggaacgtt atcttcccg aagacgggttgcg tggcacatcg ggaatcgccg
 36421 cgggtgcgtc ttgcgtcgcc agcccgagat agatccggcc cagctgtcg ggcacatggcga
 36481 acacgaagcg gttcagatca cccgggatcg gaatggtcag cgcggcagtc ggattggcaa
 36541 acgacttcgc gtcgaagagacc agatgtgtgc cgcggctggg gcgtcgatcc agggacgggt
 25 36601 cgcgttcacc cgcacccacgc cccggccgt tgcgtcgcc acgcgcacgc acgcgcacgc
 36661 actgcgggtt ggcgggtcg tcactccca cgcgtcgatcg ggtgacatcc gacgcggccca
 36721 cgtaaatgttgc gatgcggggcg cgcgtcgatcg cgcgtcgatcg cgcgtcgatcg atgaccagcc
 36781 gggcgtcgatcg gatcaatttcgc cgcgtcgatcg cgcgtcgatcg cgcgtcgatcg cgcgtcgatcg
 36841 gaacgggtggg agcaatctcc accacccgtt acgcgggatcg tggcgccatcg cggggcaacg
 36901 tgcgtcgcccg cgtacccgtt acgcacccgtt aagcgtcgcc ggcggccatcg cggccacgc
 36961 ccaaccccg ctgggtgtga cccatcgacgc gcaacaaacgg gaccagtgc ggcacggcat
 37021 gcacgagatcg aggagcgatcg cgtgtatcgatcg ggattccgcg ttcgacggcg ctgcgtcgcc
 37081 cgtggccac gttggccgtt ggcgtcgatcg cgcgtcgatcg cgcgtcgatcg ttcgacggcg
 37141 agcggctggt ggcgtcgatcg agatcatgtt ttccaccaaa ggcgtcgatcg agaccgcggg
 37201 tggcgtcgatcg taaggcaatcg ccaacaccggg taatgcgcgcg cgcgtcgatcg atgcgtcgatcg
 35 37261 gtgcgtcgatcg tggcgtcgatcg cgcgtcgatcg cgcgtcgatcg cgcgtcgatcg tggcgtcgatcg

37321 ccgagtgggg catcagcaca aatatccgtt cagtgctgg gtaagttcgg tggccagcgc
 37381 ggccggatcg aggtatcgaaat cgacgatgtc cgccggactgg atggtcact gggcgatcag
 37441 caacaccatg gtcgcccagt gacgagcgtc gccggagcgc acactgcccc accgctgcgc
 5 37501 cactgtcagc cggccggcca accccctcgat caggacctgc tggctggtgc cgaggcgctc
 37561 ggtgtatgtac accctggcca gtcggagtg catgaccgac atgatcagat cgtcaccgg
 37621 caaccggctcg gccaccgcga caatctgcgtt taccaacgtc tcccggtcg ccccgatcg
 37681 gggcaccctcc cgcagcacgt cgccgatatg gtcggatcg atggacgcca tgatecgacc
 37741 ggtgtccggc cagcgacggt atacggtcgg gccggctcagc cccgcgcgc gggcgatctc
 10 37801 ggcaagtgtc accccgttcca cgcgttaatc gacgacgcag ctgcggctg ccccgaggat
 37861 acgaccaccg gttatccgcgc ggtcattact cattgacagc atgtgtaaata ctgtaaacgcg
 37921 tgactcaccg cgaggaactc cttccaccga tgaatggga cgcgtgggaa gatcccgccg
 37981 cggccaagcc actttctgtat ggcgtccggc cgttgctgaa gcagggttg ggcctagcgg
 38041 actcggagca gcccgaactc gaccccgccg aggtgcagct ggcggccgtcc gcccgtcgg
 15 38101 gggcagacca (SEQ ID NO: 24)

5.9. X-linked Inhibitor of Apoptosis Protein (“XIAP”)

GenBank Accession # U45880:

1 gaaaagggtgg acaagtctta tttcaagag aagatgacattt ttaacagttt tgaaggatct
 20 61 aaaacttgg tacctgcaga catcaataag gaagaagaat ttgtagaaga gtttaataga
 121 taaaaaaactt ttgcttaattt tccaaatgtt agtcctgttt cagcatcaac actggcacga
 181 gcagggtttc ttatactgg tgaaggagat accgtgcggt gctttatgtt tcatgcagct
 241 gtagatagat ggcaaatatgg agactcagca gttggaaagac acaggaaagt atccccaaat
 301 tgcagattt tcaacggctt ttatcttggaa aatagtgcctt cgcagtctac aaattctggt
 361 atccagaatg gtcagttacaa agtggaaaac tatctggaa gcaagatca tttgcctta
 421 gacaggccat ctgagacaca tgcagactat cttttggaaa ctgggcagggt tgttagatata
 481 tcaagacacca tatacccgag gaaccctgccc atgtattgtt aagaagcttag attaaagttcc
 541 tttcagaact ggcggacta tgctcaccta accccaagag agtttagcaag tgctggactc
 601 tactacacag gtattgggtga ccaagtgcag tgctttgtt gtgggtggaaa actgaaaaat
 661 tgggaacattt tgatctgtc ctggcagaa cacaggcgcac actttctaa ttgcctttt
 721 gtttggggcc ggaatcttaa tattcgaagt gaatctgtat ctgtgagttc tgataggaat
 781 ttcccaattt caacaaatct tccaaagaaat ccatccatgg cagattatga agcacggatc
 841 ttatcttttggacatggat atactcagtt aacaaggagc agcttgcaag agctggattt
 901 tatgccttag gtgaaggtga taaagtaaag tgctttcact gtggaggagg gcttaactgtat
 961 tggaaagccca gtgaagaccc ttggaaacaa catgctaaat ggtatccagg gtgcaaatat
 35 1021 ctgttagaaac agaagggaca agaataatata aacaatattc attaactca ttcaacttgag

1081 gagtgtctgg taagaactac tgagaaaaca ccatcactaa ctagaagaat tgatgatacc
 1141 atcttccaaa atcctatggt acaagaagct atacgaatgg ggttcagttt caaggacatt
 1201 aagaaaataa tggagggaaaa aattcagata tctgggagca actataaatac acttgaggtt
 5 1261 ctgggtcag atcttagtcaa tgctcagaaa gacagtatgc aagatgagtc aagtcagact
 1321 tcattacaga aagagatgg tactgaagag cagctaaggc gcctgcaaga ggagaagett
 1381 tgcaaaaatct gtatggatag aaatattgct atcggttttg ttccctgtgg acatctagtc
 1441 acttgtaaac aatgtctga agcagttgac aagtgccca tgitctacac agtcattact
 1501 ttcaagcaaa aaatttttat gtcttaatct aactctatag taggcattgtt atgttgttct
 10 1561 tattaccctg attgaatgtg tgatgtgaac tgactttaag taatcaggat tgaattccat
 1621 tagcattgc taccaagtag gaaaaaaaaat gtacatggca gtgttttagt tggcaatata
 1681 atcttgaat ttcttgattt ttcaagggtat tagctgtatt atccattttt ttactgtta
 1741 tttaatgtaa accatagact aagaataaga agcatcatac tataactgaa cacaatgtt
 1801 attcatagta tactgattt atttctaagt gtaagtgaat taatcatctg gatttttat
 1861 tcttttcaga taggcttaac aaatggagct ttctgtatataaaatgtggag attagagttt
 1921 atctccccaa tcacataatt tgttttgtt gaaaaaggaa taaatgttc catgctggtg
 1981 gaaagataga gattgtttt agagggtgg tttttttt taggattctg tccattttct
 2041 tgtaaaggga taaacacggc cgtgtgcgaa atatgtttgtt aaagtgtttt gccattgtt
 2101 aaagcgattt taatgtataga atactatcga gccaacatgtt actgacatgg aaagatgtca
 2161 gagatatgtt aagtgtaaaa tgcaagtggc gggacactat gtatagtcg agccagatca
 2221 aagtatgtat ttgtttaata tgcatagaac gagagattttt gaaagatata caccaaaactg
 2281 tttaatgtgg ttctcttcg gggagggggg gattgggggg ggggccccag aggggttt
 2341 gagggccctt ttcaacttcg actttttca ttgttctg ttccggatttt ttataagtat
 2401 gtagaccccg aagggtttt tggactaa catcagtaac ctaacccccc tgactatcct
 2461 gtgccttcc tagggagctg tttttttcc caccaccac cttccctct gaacaaatgc
 25 2521 ctgagtgctg gggactttt (SEQ ID NO: 25)

General Target Region:

Internal Ribosome Entry Site (IRES) in 5' untranslated region:
 30 5'AGCUCCUAUAACAAAAGUCUGUUGCUUGUGUUUCACAUUUUGGAUUU
 CCUAAUAUAUAUGUUCUCUUUUUAGAAAAGGUGGGACAAGUCCUAUUUUC
 AAGAGAAG3' (SEQ ID NO: 26)

Initial Specific Target Motif:

35 RNP core binding site within XIAP IRES
 5'GGAUUUCCUAUAUAUGUUCUCUUUUU3' (SEQ ID NO: 27)

5.10. Survivin

GenBank Accession # NM 001168:

1 ccggccaggatt tgaatcgccgg gaccgggttgg cagaggtggc ggccggccgc tgggtgcccc
5 61 gacgttgcgg cctgcctggc agcccttct caaggaccac cgcacatctca cattcaaga
121 ctggcccttc ttggagggct gcgcctgcac cccggagcgg atggccgagg ctggctcat
181 ccactgcccc actgagaacg agccagactt ggcccagtgt ttcttcgtctcaaggagct
241 ggaaggctgg gagccagatg acgaccccat agaggaacat aaaaaggatt cgtccgggtg
301 cgcttcctt tctgtcaaga agcagttga agaattaacc ctggtaaat ttgtaaact
10 361 ggacagagaa agagccaaga acaaaaattgc aaaggaaacc aacaataaga agaaagaatt
421 tgagggaaact gcgaagaaag tgcgcgtgc catcgagcag ctggctgcca tggattgagg
481 cctctggccg gagctgcctg gtccagagt ggctgcacca ctggcagggt ttatccctg
541 gtgccaccag ctttcgtgtt ggcccttagt caatgttta ggaaaggaga tcaacattt
601 caaatttagat tttcaactg tgcttcgtt ttgtttttaa agtggcacca gaggtgcctc
15 661 tgcctgtca gcgggtgtc ctggtaacag tggctgcctc tctctctc tctcttttt
721 gggggctcat ttttgcgtt ttgattcccg ggcttaccag gtgagaagtg agggaggaag
781 aaggcagtgt ccctttgtc agagctgaca gcttgcgtc cgtggcaga gcctccaca
841 gtgaatgtt ctggaccctca ttttttttttggacttgcag ttttttttttggacttggca
901 ggtgcctgtt gaatctgagc tgcagggttcc ttatctgtca caccctgtgcc tccctcaggagg
961 acagttttt ttttttttttggacttgcag ttttttttttggacttgcag ttttttttttggacttggca
20 1021 gtgtatgagag aatggagaca ggtccctgg ctccctact gtttaacaac atggctttct
1081 ttttttttttggacttgcag ttttttttttggacttgcag ttttttttttggacttggca
1141 agccattctca agtcatggg gaaacggggtaaacttcagg tggatgagga gacagaatag
1201 agtgtatgatggc agcgctgtgc agatactcttgc ttttttttttggacttgcag ttttttttttggacttggca
1261 agtggccgc gggccacatg ctggccgtc ctccctcaga aaaaggcagt ggcctaaatc
1321 cttttttttttggacttgcag ttttttttttggacttgcag ttttttttttggacttgcag ttttttttttggacttggca
1381 tctgtcagcc caacccctcac atctgtcagc ttttttttttggacttgcag ttttttttttggacttggca
1441 ccagggtcccc gcttttttggacttgcag ttttttttttggacttgcag ttttttttttggacttggca
1501 gatggatttgc ttttttttttggacttgcag ttttttttttggacttgcag ttttttttttggacttggca
1561 gctggaaacc ttttttttttggacttgcag ttttttttttggacttgcag ttttttttttggacttggca (SEQ ID NO: 28)

The present invention is not to be limited in scope by the specific embodiments described herein. Indeed, various modifications of the invention in addition to those described will become apparent to those skilled in the art from the foregoing description and accompanying figures. Such modifications are intended to fall within the scope of the appended claims.

Various publications are cited herein, the disclosures of which are incorporated by reference in their entireties.

5

10

15

20

25

30

35

The invention can be illustrated by the following embodiments enumerated in the numbered paragraphs that follow:

- 5 1. A method for identifying a test compound that binds to a target RNA molecule, comprising the steps of (a) contacting a detectably labeled target RNA molecule with a library of solid support-attached test compounds under conditions that permit direct binding of the labeled target RNA to a member of the library of solid support-attached test compounds so that a detectably labeled target RNA:support-attached test compound complex is formed; (b) separating the detectably labeled target RNA:support-attached test compound complex formed in step (a) from uncomplexed target RNA molecules and test compounds, and (c) determining a structure of the test compound of the RNA:support-attached test compound complex.
- 10 15 2. The method of paragraph 1 in which the target RNA molecule contains an HIV TAR element, internal ribosome entry site, "slippery site", instability element, or adenylate uridylate-rich element.
- 15 20 25 3. The method of paragraph 1 in which the RNA molecule is an element derived from the mRNA for is tumor necrosis factor alpha ("TNF- α "), granulocyte-macrophage colony stimulating factor ("GM-CSF"), interleukin 2 ("IL-2"), interleukin 6 ("IL-6"), vascular endothelial growth factor ("VEGF"), human immunodeficiency virus I ("HIV-1"), hepatitis C virus ("HCV" - genotypes 1a & 1b), ribonuclease P RNA ("RNaseP"), X-linked inhibitor of apoptosis protein ("XIAP"), or survivin.
- 30 35 4. The method of paragraph 1 in which the detectably labeled RNA is labeled with a fluorescent dye, phosphorescent dye, ultraviolet dye, infrared dye, visible dye, radiolabel, enzyme, spectroscopic colorimetric label, affinity tag, or nanoparticle.
- 30 35 5. The method of paragraph 1 in which the test compound is selected from a combinatorial library comprising peptoids; random bio-oligomers; diversomers such as hydantoins, benzodiazepines and dipeptides; vinylogous polypeptides; nonpeptidal peptidomimetics; oligocarbamates; peptidyl phosphonates; peptide nucleic acid libraries; antibody libraries; carbohydrate libraries; and small organic molecule libraries including, but not limited to, benzodiazepines, isoprenoids, thiazolidinones, metathiazanones, pyrrolidines, morpholino compounds, or diazepindiones.

6. The method of paragraph 1 in which screening a library of test
compounds preferably comprises contacting the test compound with the target nucleic acid in
5 the presence of an aqueous solution, the aqueous solution comprising a buffer and a
combination of salts, preferably approximating or mimicking physiologic conditions

7. The method of paragraph 6 in which the aqueous solution optionally
further comprises non-specific nucleic acids comprising DNA, yeast tRNA, salmon sperm
10 DNA, homoribopolymers, and nonspecific RNA.

8. The method of paragraph 6 in which the aqueous solution further
comprises a buffer, a combination of salts, and optionally, a detergent or a surfactant. In
another embodiment, the aqueous solution further comprises a combination of salts, from
15 about 0 mM to about 100 mM KCl, from about 0 mM to about 1 M NaCl, and from about 0
mM to about 200 mM MgCl₂. In a preferred embodiment, the combination of salts is about
100 mM KCl, 500 mM NaCl, and 10 mM MgCl₂. In another embodiment, the solution
optionally comprises from about 0.01% to about 0.5% (w/v) of a detergent or a surfactant.

9. Any method that detects an altered physical property of a target
nucleic acid complexed to a test compound attached to a solid support from the unbound
target nucleic acid may be used for separation of the complexed and non-complexed target
nucleic acids in the method of paragraph 1. Methods such as flow cytometry, affinity
chromatography, manual batch mode separation, suspension of beads in electric fields, and
25 microwave are used for the separation of the complexed and non-complexed target nucleic
acids.

10. The structure of the substantially one type of test compound of the
RNA:test compound complex of paragraph 1 is determined, in part, by the type of library of
30 test compounds. In a preferred embodiment wherein the combinatorial libraries are small
organic molecule libraries, mass spectroscopy, NMR, or vibration spectroscopy are used to
determine the structure of the test compounds. In an embodiment wherein the combinatorial
libraries are peptide or peptide-based libraries, Edman degradation is used to determine the
structure of the test compounds.

WHAT IS CLAIMED IS:

1. A method for identifying a test compound that binds to a target RNA molecule, comprising the steps of:
 - 5 (a) contacting a detectably labeled target RNA molecule with a library of solid support-attached test compounds under conditions that permit direct binding of the labeled target RNA to a member of the library of solid support-attached test compounds so that a detectably labeled target RNA:support-attached test compound complex is formed;
 - (b) separating the detectably labeled target RNA:support-attached test compound complex formed in step (a) from uncomplexed target RNA molecules and test compounds by flow cytometry; and
 - 10 (c) determining a structure of the substantially one type of test compound of the RNA:support-attached test compound complex by mass spectroscopy.

20

25

30

35

SEQUENCE LISTING

<110> PCT Therapeutics, Inc.

<120> METHODS FOR IDENTIFYING SMALL MOLECULES THAT BIND SPECIFIC RNA STRUCTURAL MOTIFS

<130> 10589-008-228

<140> To be assigned

<141> 2002-04-11

<150> 60/282,966

<151> 2001-04-11

<160> 28

<170> PatentIn version 3.0

<210> 1

<211> 21

<212> RNA

<213> Homo sapiens

<400> 1

auuuuauuuau uuauuuauuu a

21

<210> 2

<211> 17

<212> RNA

<213> Homo sapiens

<400> 2

auuuuauuuau uuauuuua

17

<210> 3
<211> 15
<212> RNA
<213> Homo sapiens

<400> 3
wauuuauuuua uuuuaw

15

<210> 4
<211> 13
<212> RNA
<213> Homo sapiens

<400> 4
wwauuuauuu aww

13

<210> 5
<211> 13
<212> RNA
<213> Homo sapiens

<400> 5
wwwauuuuaw www

13

<210> 6
<211> 1643
<212> DNA
<213> Homo sapiens

<400> 6
gcagaggacc agctaagagg gagagaagca actacagacc ccccctgaaa acaaccctca 60
gacgccacat cccctgacaa gctgccaggc aggttctctt cctctcacat actgacccac 120
ggctccaccc tctctccctt ggaaaggaca ccatgagcac tgaaagcatg atccgggacg 180
tggagctggc cgaggaggcg ctccccaaaga agacaggggg gccccaggc tccaggcggt 240
gcttgttcct cagcctcttc tccttcctga tcgtggcagg cgccaccacg ctcttctgcc 300

| | |
|---|------|
| tgctgactt tggagtgatc ggcccccaga gggaaagagtt ccccaggac ctctctctaa | 360 |
| tcagccctct ggcccaggca gtcagatcat cttctcgaac cccgagtgac aagcctgtag | 420 |
| cccatgttgt agcaaaccct caagctgagg ggcagctcca gtggctgaac cgccgggcca | 480 |
| atgccttcct ggccaatggc gtggagctga gagataacca gctggtggtg ccatcagagg | 540 |
| gcctgtacct catctactcc caggtcotct tcaagggcca aggctgcccc tccacccatg | 600 |
| tgctcctcac ccacaccatc agccgcacatcg ccgtctccta ccagaccaag gtcaacctcc | 660 |
| tctctgccat caagagcccc tgccagaggg agacccaga gggggctgag gccaagccct | 720 |
| ggtatgagcc catctatctg ggaggggtct tccagctgga gaagggtgac cgactcagcg | 780 |
| ctgagatcaa tcggcccgac tatctcgact ttgccgagtc tggcaggtc tactttggaa | 840 |
| tcattgccct gtgaggagga cgaacatcca accttccaa acgcctcccc tgccccaatc | 900 |
| cctttattac cccctcccttc agacaccctc aacctcttct ggctaaaaa gagaattggg | 960 |
| ggcttagggt cgaaacccaa gcttagaact ttaagcaaca agaccaccac ttcgaaacct | 1020 |
| gggattcagg aatgtgtggc ctgcacagtg aattgctggc aaccactaag aattcaaact | 1080 |
| ggggcctcca gaactcaactg gggcctacag ctttgcatccc tgacatctgg aatctggaga | 1140 |
| ccagggagcc tttggttctg gccagaatgc tgcaggactt gagaagacct cacctagaaa | 1200 |
| ttgacacaag tggacccttag gccttcctct ctccagatgt ttccagactt ctttgagaca | 1260 |
| cggagccctag ccctccccat ggagccagct ccctctattt atgtttgcac ttgtgattat | 1320 |
| ttattattta ttattatttt atttattttc agatgaatgt atttattttgg gagaccgggg | 1380 |
| tatcctgggg gacccaatgt aggagctgcc ttggctcaga catgtttcc gtgaaaacgg | 1440 |
| agctgaacaa taggctgttc ccatgttagcc ccctggccctc tgccttct ttgattatg | 1500 |
| ttttttaaaa tatttatctg attaagttgt ctaaacaatg ctgatttgg gaccaactgt | 1560 |
| cactcattgc tgagcctctg ctccccaggg gagttgtgtc tgtaatcgcc ctactattca | 1620 |
| gtggcgagaa ataaagtttgc ctt | 1643 |

<210> 7

<211> 756

<212> DNA

<213> Homo sapiens

| | |
|--|-----|
| <400> 7 | |
| gctggaggat gtggctgcag agcctgcgtc tcttggcac tgtggcctgc agcatctctg | 60 |
| cacccgccccg ctgcggcagc cccagcacgc agccctggga gcatgtgaat gccatccagg | 120 |
| aggcccccggc tctcctgaac ctgagtagag acactgctgc tgagatgaat gaaacagtag | 180 |

aagtcatctc agaaatgttt gacctccagg agccgacctg cctacagacc cgcctggagc 240
 tgtacaagca gggcctgcgg ggcagcctca ccaagctcaa gggccccttg accatgatgg 300
 ccagccacta caagcagcac tgccctccaa ccccgaaac ttccctgtgca acccagacta 360
 tcaccttga aagtttcaaa gagaacctga aggactttct gcttgcata cccttgact 420
 gctggagcc agtccaggag tgagaccggc cagatgaggc tggccaagcc ggggagctgc 480
 tctctcatga aacaagagct agaaactcag gatggcata tcggagggac caaggggtgg 540
 gccacagcca tggggaggt ggcctggacc tgccctgggc cacactgacc ctgatacagg 600
 catggcagaa gaatgggaat attttatact gacagaaatc agtaatattt atatatttat 660
 atttttaaaa tattttatata tttatattt taagttcata ttccatattt attcaagatg 720
 ttttaccgta ataatttata taaaaatat gcttct 756

<210> 8

<211> 756

<212> DNA

<213> Homo sapiens

<400> 8
 tctggaggat gtggctgcag agcctgctgc tcttggcac tggcctgc agcatctcg 60
 caccggcccg ctgcggccagc cccagcacgc agccctggga gcatgtaat ggcacccagg 120
 aggccggcg tctcctgaac ctgagtagag acactgctgc tgagatgaat gaaacagtag 180
 aagtcatctc agaaatgttt gacctccagg agccgacctg cctacagacc cgcctggagc 240
 tgtacaagca gggcctgcgg ggcagcctca ccaagctcaa gggccccttg accatgatgg 300
 ccagccacta caagcagcac tgccctccaa ccccgaaac ttccctgtgca acccagacta 360
 tcaccttga aagtttcaaa gagaacctga aggactttct gcttgcata cccttgact 420
 gctggagcc agtccaggag tgagaccggc cagatgaggc tggccaagcc ggggagctgc 480
 tctctcatga aacaagagct agaaactcag gatggcata tcggagggac caaggggtgg 540
 gccacagcca tggggaggt ggcctggacc tgccctgggc cacactgacc ctgatacagg 600
 catggcagaa gaatgggaat attttatact gacagaaatc agtaatattt atatatttat 660
 atttttaaaa tattttatata tttatattt taagttcata ttccatattt attcaagatg 720
 ttttaccgta ataatttata taaaaatat gcttct 756

<210> 9

<211> 825

<212> DNA

<213> Homo sapiens

<400> 9
atcactctct ttaatcacta ctcacattaa cctcaactcc tgccacaatg tacaggatgc 60
aactcctgtc ttgcattgca ctaattctt cacttgcac aaacagtgc a cctacttcaa 120
gttcgacaaa gaaaacaaaag aaaacacagc tacaactgga gcatttactg ctggatttac 180
agatgatttt gaatggaatt aataattaca agaatccaa actcaccagg atgctcacat 240
ttaagttta catgcccag aaggccacag aactgaaaca gcttcagtgt ctagaagaag 300
aactcaaacc tctggaggaa gtgctgaatt tagctcaaag caaaaacttt cacttaagac 360
ccagggactt aatcagcaat atcaacgtaa tagttctgga actaaaggga tctgaaacaa 420
cattcatgtg tgaatatgca gatgagacag caaccattgt agaatttctg aacagatgga 480
ttaccttttgc tcaaagcatc atctcaacac taacttgata attaagtgt tcccacttaa 540
aacatatcag gccttctatt tatttatTTT aatattttaa ttTTtatTTT attgttgaat 600
gtatggttgc tacctattgt aactattatt cttaatctta aaactataaa tatggatctt 660
ttatgattct ttTgttaagc cctagggct ctaaaatggt ttaccttatt tatccaaaaa 720
atatttatta ttatgttcaa tgTTaaatTTT agtatctatg tagattggTTT agtaaaacta 780
tttaataaaat ttgataaataa taaaaaaaaaaa aaacaaaaaaaaaa aaaaa 825

<210> 10

<211> 15

<212> RNA

<213> Homo sapiens

<220>

<221> misc_feature

<222> (1)..(1)

<223> N = A, U, G, OR C

<220>

<221> misc_feature

<222> (15)..(15)

<223> N = A, U, G, OR C

<400> 10
nauuuauuuua uuuuau 15

<210> 11

<211> 1125

<212> DNA

<213> Homo sapiens

<400> 11
ttctgcctc gagcccaccc ggaacgaaag agaacgtcta tctgcctcc aggagccag 60
ctatgaactc cttctccaca agcgccttcg gtccagttgc cttctccctg gggctgctcc 120
tggtgttgcc tgctgccttc cctgccccag tacccccagg agaagattcc aaagatgttag 180
ccgccccaca cagacagcca ctcacccctt cagaacgaat tgacaaacaa attcggtaca 240
tcctcgacgg catctcgacc ctgagaaagg agacatgtaa caagagtaac atgtgtgaaa 300
gcagcaaaga ggcactggca gaaaacaacc tgaaccttcc aaagatggct gaaaaagatg 360
gatgcttcca atctggattc aatgaggaga cttgccttgt gaaaatcatc actggcttt 420
tggagtttga ggtataccta gagtacctcc agaacagatt tgagagtagt gaggaacaag 480
ccagagctgt gcagatgagt acaaaagtcc tgatccagtt cctgcagaaa aaggcaaaga 540
atctagatgc aataaccacc cctgacccaa ccacaaatgc cagcctgctg acgaagctgc 600
aggcacagaa ccagtggctg caggacatga caactcatct cattctgcgc agctttaagg 660
agttcctgca gtccagcctg agggctttc ggcaaatgta gcatgggcac ctcagattgt 720
tgttgttaat gggcattcct tcttctggtc agaaacctgt ccactggca cagaacttat 780
gttggttctct atggagaact aaaagtatga gcgttaggac actattttaa ttattttaa 840
tttattaata tttaaatatg tgaagctgag ttaatttatg taagtcatat ttatatttt 900
aagaagtacc acttggaaaca ttttatgtat tagtttgaa ataataatgg aaagtggcta 960
tgcagttga atatccttgc ttccagagcc agatcatttc ttggaaagtg taggcttacc 1020
tcaaataaaat ggctaactta tacatatttt taaagaaata ttatattgt atttatataa 1080
tgtataaaatg gttttatac caataaaatgg cattttaaaa aattc 1125

<210> 12

<211> 3166

<212> DNA

<213> Homo sapiens

<400> 12
 aagagctcca gagagaagtc gaggaagaga gagacggggt cagagagac gcgcggcgt 60
 gcgagcagcg aaagcgacag gggcaaagtg agtacactgc ttttgggggt gaccggcga 120
 gcgccggcgtg agccctcccc cttggatcc cgcaactgac cagtcgcgt gacggacaga 180
 cagacagaca ccgcggccag ccccaagttac cacccctcc cccggccggcg gcggacagt 240
 gacgcggcgg cgagccgcgg gcagggggcg gagccgcggc ccggaggcgg ggtggagggg 300
 gtcggagctc gcggcgtcgc actgaaactt ttctccaac ttctgggctg ttctcgcttc 360
 ggaggagccg tggtccgcgc gggggaaagcc gagccgagcg gagccgcgag aagtgcgtac 420
 tcgggcggg aggagccgca gcgggaggag ggggaggagg aagaagagaa ggaagaggag 480
 agggggccgc agtggcgact cggcgctcgg aagccggct catggacggg tgaggcggcg 540
 gtgtgcgcag acagtgcgtcc acgcgcgcgc ctcccaagcc ctggcccgcc ctcggccgg 600
 gaggaagagt agctcgccga ggcgcccggg agagcgggccc gcggcacagc ccgagccgga 660
 gagggacgcgc agccgcgcgc cccggtcggg cctccgaaac catgaacttt ctgctgtctt 720
 gggtgcatgg gggcttgcc ttgctgctt acctccacca tgccaaatggg tccaggctg 780
 caccatggc agaaggagga gggcagaatc atcacgaatg ggtgaatgtc atggatgtct 840
 atcagcgcag ctactgccat ccaatcgaga ccctggtgga catcttccag gagtaccctg 900
 atgagatcga gtacatcttc aagccatcct gtgtccctt gatgcgtatgc ggggctgtct 960
 ccaatgacga gggctggag tttgtccca ctgaggagtc caacatcacc atgcagatta 1020
 tgccggatcaa acctcacca ggcgcggcaca taggagagat gagttcccta cagcacaaca 1080
 aatgtgaatg cagaccaaag aaagatagag caagacaaga aaatccctgt gggccttgct 1140
 cagagcggag aaagcattt tttgtacaag atccgcagac gtgtaaatgt tccatgggg 1200
 acacacactc gcgttgcag ggcaggcagc ttgagttaaa cgaacgtact tgcagatgt 1260
 acaagccgag gcggtgagcc gggcaggagg aaggagcctc cctcagggtt tcgggaaacca 1320
 gatctctctc cagggaaagac tgatacagaa cgtatcgatac agaaaccacg ctgcgcac 1380
 cacaccatca ccatcgacag aacagtccctt aatccagaaa cctgaaatga aggaagagga 1440
 gactctgcgc agagcacttt gggtccggag ggcgcggactc cggcggaaacgc atccccggc 1500
 gggtgaccca gcacggtccc tcttggaaatt ggattcgcca ttttattttt cttgctgtca 1560
 aatcaccgag cccggaagat tagagagttt tatttctggg attcctgttag acacacccac 1620
 ccacatacat acatttatat atatatatat tatatatata taaaaataaa tatctctatt 1680
 ttatatatat aaaatatata tattttttt ttaaattaac agtgcataatg ttattgggt 1740
 cttcactgga ttttttgc ttttttttgc ttgagttggg agggaaatgt tcccaactcag 1800

| | |
|---|------|
| atcctgacag ggaagaggag gagatgagag actctggcat gatcttttt ttgtcccact | 1860 |
| tggtggggcc agggtcctct cccctgccc agaardtgca aggccaggc atggggcaa | 1920 |
| atatgaccca gtttggaa caccgacaaa cccagccctg gcgctgagcc tctctacccc | 1980 |
| aggtcagacg gacagaaaga caaatcacag gttccggat gaggacaccc gctctgacca | 2040 |
| ggagttggg gagcttcagg acattgtgt gcttggga ttccctccac atgctgcacg | 2100 |
| cgcacatctgc cccaggggc actgcctgga agattcagga gcctggcgg cttcgctta | 2160 |
| ctctcacctg ctctcgatgt gcccaggagg ccactggcag atgtcccgcc gaagagaaga | 2220 |
| gacacattgt tggagaaggc agcccatgac agcgcctt cctggactc gccctcatcc | 2280 |
| tcttcctgct ccccttcctg gggtgcagcc taaaaggacc tatgtcctca caccattgaa | 2340 |
| accactagtt ctgtccccc agaaacctg gttgtgtgt tggatgtgt tgaccttcct | 2400 |
| ccatccctg gtccttcct tccctcccg aggacacagag agacaggcga ggatccacgt | 2460 |
| gcccattgtg gaggcagaga aaagagaaag tggatgtat acggtaactt ttaatatcc | 2520 |
| cttttaatt agaaattaga acagtttaatt taattaaaga gtagggttt tttcagttat | 2580 |
| tcttggtaa tatttaattt caactattt tgagatgtat ctttgcctct ctctgcct | 2640 |
| cttatttgc ccggtttttgc tatataaaat tcatgttcc aatctctctc tccctgatcg | 2700 |
| gtgacagtca ctgcatttgc ttgaacagat atttaattt gctaacaactc agctctgccc | 2760 |
| tccccgatcc cctggctccc cagcacacat tccttgaaa gagggttca atatacatct | 2820 |
| acataactata tatataattgg gcaacttgc tttgtgtgtat tatataatata tatatgtta | 2880 |
| tgtatataatg tgatcctgaa aaaataaaca tcgcatttct gtttttata tggtaaaacc | 2940 |
| aaacaagaaa aaatagagaa ttctacatc taaatctctc tcctttttt atttaatata | 3000 |
| ttgttatcat ttatttatttgc tgctactgt ttatccgtaa taattgtgg gaaaagatata | 3060 |
| taacatcacg tcttgcctc tagtgcagtt ttccgagata ttccgtagta catattttt | 3120 |
| tttaaacaac gacaaagaaa tacagatata tcttaaaaaa aaaaaaa | 3166 |

<210> 13

<211> 249

<212> RNA

<213> Homo sapiens

| | |
|--|-----|
| <400> 13 | |
| ccgggcuau ggacggugaa ggcggcggug ugcgcagaca gugcuccagc ggcgcgcuc | 60 |
| cccagccug gccccggccuc gggccgggag gaagaguagc ugcggagggc gcccggaga | 120 |
| gcggggccgc ccacagcccg agccggagag ggacgcgagc cgccgcgcgc ggucggccu | 180 |

ccgaaaccau gaacuuucug cugucuuggg ugcauuggag cnuugccuug cugcucuacc 240
 uccaccaug 249

<210> 14

<211> 9181

<212> DNA

<213> *Homo sapiens*

<400> 14
 ggtctctctg gttagaccag atctgaggct gggagctctc tggcttaacta gggaaacccac 60
 tgcttaagcc tcaataaaagc ttgcctttag tgcctcaagt agtgtgtgcc cgtctgttgt 120
 gtgactctgg taactagaga tccctcagac ccttttagtc agtgtggaaa atctctagca 180
 gtggcgccccg aacagggacc tgaaagcgaa agggaaacca gaggagctct ctcgacgcag 240
 gactcggctt gctgaagcgc gcacggcaag aggcgagggg cggcgactgg tgagtacgcc 300
 aaaaattttg actagcggag gctagaagga gagagatggg tgcgagagcg tcagtattaa 360
 gcgggggaga attagatcga tggaaaaaaa ttcgggttaag gccaggggg aagaaaaaat 420
 ataaattaaa acatatagtt tggcaagca gggagctaga acgattcgcg gttaatcctg 480
 gcctgtttaga aacatcagaa ggctgttagac aaatactggg acagctacaa ccatcccttc 540
 agacaggatc agaagaactt agatcattat ataatacagt agcaaccctc tattgtgtgc 600
 atcaaaggat agagataaaa gacaccaagg aagctttaga caagatagag gaagagcaaa 660
 acaaaaagtaa gaaaaaagca cagcaagcag cagctgacac aggacacacgc aatcaggta 720
 gccaaaattta ccctatagtg cagaacatcc agggcaaat ggtacatcag gccatatac 780
 ctagaacttt aaatgcattt gtaaaagttag tagaagagaa ggctttcagc ccagaagtga 840
 tacccatgtt ttcagcatta tcagaaggag ccacccaca agatttaaac accatgctaa 900
 acacagtggg gggacatcaa gcagccatgc aaatgttaaa agagaccatc aatgaggaag 960
 ctgcagaatg ggatagagtg catccagtgc atgcaggcc tattgcacca ggccagatga 1020
 gagaaccaag gggaaagtgc atagcaggaa ctactagtac cttcaggaa caaataggat 1080
 ggatgacaaa taatccaccc atcccagttag gagaattta taaaagatgg ataatcctgg 1140
 gattaaataa aatagtaaga atgtatagcc ctaccagcat tctggacata agacaaggac 1200
 caaaggaacc ctttagagac tatgttagacc ggttctataa aactctaaga gccgagcaag 1260
 cttcacagga ggtaaaaat tggatgacag aaaccttgtt ggtccaaaat gcgaacccag 1320
 attgttaagac tattttaaaa gcattggac cagcggctac actagaagaa atgtacac 1380
 catgtcaggg agtaggagga cccggccata aggcaagagt tttggctgaa gcaatgagcc 1440

aagtaacaaa ttcaagctacc ataatgatgc agagaggcaa ttttaggaac caaagaaaaga 1500
ttgttaagt tttcaattgt ggcaaagaag ggcacacagc cagaaattgc agggccctta 1560
ggaaaaaggg ctgttgaaa tgtggaaagg aaggacacca aatgaaagat tgtactgaga 1620
gacaggctaa ttttttaggg aagatctggc ctccctacaa gggaaaggcca gggaaattttc 1680
ttcagagcag accagagcca acagccccac cagaagagag cttcaggctc ggggttagaga 1740
caacaactcc ccctcagaag caggagccga tagacaagga actgtatcct ttaacttccc 1800
tcaggtcact ctttggcaac gacccctcgta cacaataaag atagggggc aactaaagga 1860
agctctatta gatacaggag cagatgatac agtattagaa gaaatgagtt tgccaggaag 1920
atggaaacca aaaatgatac gggaaattgg aggttttatac aaagtaagac agtatgatca 1980
gatactcata gaaatctgtg gacataaagc tataggtaca gtattagtag gacctacacc 2040
tgtcaacata attggaagaa atctgttgc tcagattggc tgcaactttaa atttcccat 2100
tagccctatt gagactgtac cagtaaaatt aaagccagga atggatggcc caaaagttaa 2160
acaatggcca ttgacagaag aaaaaataaa agcattagta gaaattttgtc cagagatggc 2220
aaaggaaggg aaaatttcaa aaattggcc tgaaaatcca tacaatactc cagtatttgc 2280
cataaagaaa aaagacagta ctaaatggag aaaattagta gatttcagag aacttaataa 2340
gagaactcaa gacttctggg aagttcaatt aggaataccca catcccgag gttaaaaaaa 2400
gaaaaaatca gtaacagtac tggatgtggg tggatgttgc ttgcatttc ctttagatga 2460
agacttcagg aagtatactg catttaccat acctagtata aacaatgaga caccaggat 2520
tagatatcag tacaatgtgc ttccacaggg atggaaagga tcaccagcaa tattccaaag 2580
tagcatgaca aaaatcttag agccttttag aaaacaaaat ccagacatag ttatctatca 2640
atacatggat gatttgtatg taggatctga cttagaaata gggcagcata gaacaaaaat 2700
agaggagctg agacaacate tggatgtggg gggacttacc acaccagaca aaaaacatca 2760
gaaagaacct ccattccctt ggttgggtt tggatgtggg tggatgttgc ttgcatttc ctttagatga 2820
gcctatagtg ctggccagaaa aagacagctg gactgtcaat gacatacaga agttatggg 2880
gaaattgaat tggcaagtc agatttaccc agggattaaa gtaaggcaat tatgtaaact 2940
ccttagagga accaaagcac taacagaagt aataccacta acagaagaag cagagctaga 3000
actggcagaa aacagagaga ttctaaaaga accagtacat ggagtgttatt atgaccatc 3060
aaaagactta atagcagaaa tacagaagca gggcaaggc caatggacat atcaaattta 3120
tcaagagcca tttaaaaatc tgaaaacagg aaaatatgca agaatgaggg gtggcccacac 3180
taatgtatgtaa aacaattaa cagaggcagt gcaaaaaata accacagaaa gcatagtaat 3240
atggggaaag actcctaaat ttggatgtggc cataaaaaag gaaacatggg aaacatggg 3300
gacagagttat tggcaagccca cctggattcc tgagtggqat tttgttaata cccctccctt 3360

| | | | | | | |
|-------------|-------------|------------|-------------|-------------|-------------|------|
| agtcaaattt | tggtaccagt | tagagaaaga | acccatagta | ggagcagaaa | ccttctatgt | 3420 |
| agatggggca | gctaacacagg | agactaaatt | aggaaaagca | ggatatgtt | ctaatacgagg | 3480 |
| aagacaaaaa | gttgcaccc | taactgacac | aacaaatcg | aagactgagt | tacaagcaat | 3540 |
| ttatctagct | ttgcaggatt | cgggattaga | agtaaacata | gtaacagact | cacaatatgc | 3600 |
| attaggaatc | attcaaggcac | aaccagatca | aagtgaatca | gagttagtc | atcaaataat | 3660 |
| agagcagttt | ataaaaaaagg | aaaaggtcta | tctggcatgg | gtaccagcac | acaaaggaat | 3720 |
| tggagggaaat | gaacaagtag | ataaaattgt | cagtgcgtt | atcagggaaag | tactatttt | 3780 |
| agatggaata | gataaggccc | aagatgaaca | tgagaaatat | cacagtaatt | ggagagcaat | 3840 |
| ggctagtgtat | tttaacctgc | cacctgttgt | agcaaaagaa | atagtagcca | gctgtgataa | 3900 |
| atgtcagcta | aaaggagaag | ccatgcatgg | acaagtagac | tgtgtccag | gaatatggca | 3960 |
| actagattgt | acacatttag | aaggaaaagt | tatccctggta | gcagttcatg | tagccagtgg | 4020 |
| atatatagaa | gcagaagtta | ttccagcaga | aacagggcag | gaaacagcat | attttctttt | 4080 |
| aaaattagca | ggaagatggc | cagtaaaaac | aatacatact | gacaatggca | gcaatttcac | 4140 |
| cggtgctacg | gttagggccg | cctgttgtg | ggcgggaatc | aagcaggaat | ttggaattcc | 4200 |
| ctacaatccc | caaagtcaag | gagtagtaga | atctatgaat | aaagaattaa | agaaaattat | 4260 |
| aggacaggtt | agagatcagg | ctgaacatct | taagacagca | gtacaaatgg | cagtattcat | 4320 |
| ccacaatttt | aaaagaaaag | gggggattgg | gggttacagt | gcaggggaaa | gaatagtaga | 4380 |
| cataatagca | acagacatac | aaactaaaga | attacaaaaa | caaattacaa | aaattcaaaa | 4440 |
| tttcgggtt | tattacaggg | acagcagaaa | tccactttgg | aaaggaccag | caaagctcct | 4500 |
| ctggaaaggt | gaagggggcag | tagtaataca | agataatagt | gacataaaag | tagtgc当地 | 4560 |
| aagaaaagca | aagatcatta | gggattatgg | aaaacagatg | gcaggtgtat | attgtgtggc | 4620 |
| aagtagacag | gatgaggatt | agaacatgga | aaagtttagt | aaaacaccat | atgtatgttt | 4680 |
| cagggaaagc | tagggatgg | ttttatagac | atcaactatga | aagccctcat | ccaagaataa | 4740 |
| tttcagaagt | acacatccc | ctagggatg | ctagatttgt | aataacaaca | tattgggtc | 4800 |
| tgcatacagg | agaaagagac | tggcatttgg | gtcagggagt | ctccatagaa | tggagggaaa | 4860 |
| agagatata | cacacaagta | gaccctgaac | tagcagacca | actaattcat | ctgtattact | 4920 |
| ttgactgttt | ttcagactct | gctataagaa | aggccttatt | aggacacata | gttagcccta | 4980 |
| ggtgtgaata | tcaagcagga | cataacaagg | tagatctct | acaatacttg | gcactagcag | 5040 |
| cattaataac | accaaaaaaag | ataaagccac | ctttgcctag | tgttacgaaa | ctgacagagg | 5100 |
| atagatggaa | caagccccag | aagaccaagg | gccacagagg | gagccacaca | atgaatggac | 5160 |
| actagagctt | ttagaggagc | ttaagaatga | agctgttaga | cattttccta | ggatttggt | 5220 |
| ccatggctta | ggcaacata | tctatgaaac | ttatgggat | acttggcag | gagtggaaac | 5280 |

cataataaga attctgcaac aactgctgtt tatccatttt cagaattggg tgtcgacata 5340
gcagaatagg cgttactcga cagaggagag caagaaatgg agccagtata tcctagacta 5400
gagccctgga agcatccagg aagtcagcct aaaactgctt gtaccaattt ctattgtaaa 5460
aagtgttgct ttcattgcca agtttggttc ataacaaaag ccttaggcat ctcctatggc 5520
aggaagaagc ggagacagcg acgaagagct catcagaaca gtcagactca tcaagcttct 5580
ctatcaaagc agtaagtagt acatgtaatg caacctatac caatagtagc aatagtagca 5640
tttagtagtag caataataat agcaatagtt gtgtggtcca tagtaatcat agaatatagg 5700
aaaatattaa gacaaagaaa aatagacagg ttaattgata gactaataga aagagcagaa 5760
gacagtggca atgagagtga aggagaaata tcagcacttg tggagatggg ggtggagatg 5820
gggcaccatg ctccttggga tggatgtat ctgttagtgct acagaaaaat tggggtcac 5880
agtctattat ggggtacctg tggatggaa agcaaccacc actctatttt gtgcacatcaga 5940
tgctaaagca tatgatacag aggtacataa tggggggcc acacatgcct gtgtacccac 6000
agaccccaac ccacaagaag tagtatttgtt aaatgtgaca gaaaattttt acatgtggaa 6060
aaatgacatg gtggacacaga tggatgagga tataatcagt ttatgggatc aaagcctaaa 6120
gccatgtgtttaaatttaccc cactctgtgt tagttaaag tgcactgatt tgaagaatga 6180
tactaataacc aatagtagta gccccgagaat gataatggag aaaggagaga taaaaactg 6240
ctctttcaat atcagcacaa gcataagagg taagggtcag aaagaatatg cattttttta 6300
taaacttgcataataccat tagataatga tactaccagc tataagttga caagttgtaa 6360
cacctcagtc attacacagg cctgtccaaa ggtatcctt gagccaattc ccatacattt 6420
tttgcccccg gctggttttt cgattctaaa atgtataat aagacgttca atggacagg 6480
accatgtaca aatgtcagca cagtacaatg tacacatggc attaggccag tagtatcaac 6540
tcaactgctg ttaaatggca gtctagcaga agaagaggtt gtaatttagat ctgtcaattt 6600
cacggacaat gctaaaacca taatgtaca gctgaacaca tctgtagaaaa ttaattgtac 6660
aagaccccaac aacaatacaa gaaaaagaat ccgtatccag agaggaccag ggagagcatt 6720
tgttacaata gaaaaatag gaaatatgag acaagcacat tggatcattt gtagagcaaa 6780
atggataaac actttaaaac agatagctag caaattaaga gaacaattttt gaaataataa 6840
aacaataatc tttaagcaat cctcaggagg ggacccagaa attgtaacgc acagttttaa 6900
tttgccgggg gaatttttct actgtaaattc aacacaactg tttaatagta cttggttttaa 6960
tagtacttgg agtactgaag ggtcaataa cactgaagga agtgacacaa tcaccctccc 7020
atgcagaata aaacaaatta taaacatgtg gcagaaagta gaaaaagcaa tggatggggcc 7080
tcccatcagt ggacaaatta gatgttcatc aaatattaca gggctgctat taacaagaga 7140
tggatggtaat agcaacaatg agtcccgat cttcagaccc ggaggaggag atatqaaqqqa 7200

caattggaga agtgaattat ataaatataa agtagtaaaa attgaaccat taggagtagc 7260
acccaccaag gcaaagagaa gagtggtgc gagagaaaaa agagcagtgg gaataggagc 7320
tttggcctt ggggttctgg gagcagcagg aagoactatg ggccgcagcct caatgacgct 7380
gacggtacag gccagacaat tattgtctgg tatagtgcag cagcagaaca atttgcgtgag 7440
ggctattgag ggcgaacacgc atctgttgc actcacagtc tggggcatca agcagctcca 7500
ggcaagaatc ctggctgtgg aaagataacct aaaggatcaa cagctcctgg ggatttgggg 7560
ttgctctgga aaactcattt gcaccactgc tgtgccttgg aatgcgtt ggagtaataa 7620
atctctggaa cagatttggaa atcacacgc ctggatggag tgggacagag aaattaacaa 7680
ttacacaagc ttaatacact ccttaattga agaatcgcaa aaccagcaag aaaagaatga 7740
acaagaatta ttggaatttag ataaatggc aagtttggg aattggttt acataacaaa 7800
ttggctgtgg tatataaaat tattcataat gatagtagga ggcttggtag gtttaagaat 7860
agttttgct gtactttcta tagtgaatag agttaggcag ggatattcac cattatcggt 7920
tcagacccac ctcccaaccc cgaggggacc cgacaggccc gaaggaatag aagaagaagg 7980
tggagagaga gacagagaca gatccattcg attagtgaac ggatccttgg cacttatctg 8040
ggacgatctg cggagcctgt gccttcttag ctaccaccgc ttgagagact tactcttgat 8100
tgtaacgagg attgtggAAC ttctggac caggggggtgg gaagccctca aatattggtg 8160
gaatctccta cagttttggaa gtcaggaact aaagaatagt gctgttagct tgctcaatgc 8220
cacagccata gcagtagctg aggggacaga tagggttata gaagtagtac aaggagcttg 8280
tagagctatt cgccacatac cttagaagaat aagacaggc ttggaaagga ttttgcata 8340
agatgggtgg caagtggtca aaaagttagt tgattggat gcctactgt agggaaagaa 8400
tgagacgagc tgagccagca gcagataggg tggagcagc atctcgagac ctggaaaaac 8460
atggagcaat cacaagtagc aatacagcag ctaccaatgc tgcttgcctt tggctagaag 8520
cacaagagga ggaggaggtg gttttccag tcacacctca ggtaccttta agaccaatga 8580
cttacaaggc agctgttagat cttagccact tttaaaaga aaagggggga ctggaaaggc 8640
taattcactc ccaaagaaga caagatatcc ttgatctgtg gatctaccac acacaaggct 8700
acttccctga ttagcagaac tacacaccag ggccagggtt cagatatcca ctgacccttg 8760
gatggtgcta caagcttagta ccagttggc cagataagat agaagaggcc aataaaggag 8820
agaacaccag ctgttacac cctgtgagcc tgcatggat ggatgaccgc gagagagaag 8880
tgtagagtg gaggtttgac agccgcctag catttcatca cgtggcccgagctgcata 8940
cgaggactt caagaactgc tgacatcgag cttgtacaa gggactttcc gctggggact 9000
ttccaggag gcgtggcctg ggcgggactg gggagtggcg agccctcaga tcctgcata 9060
aagcagctgc tttttgcctg tactgggtct ctctggtag accagatctg agcctggag 9120

ctctctggct aactaggaa cccactgctt aagcctcaat aaagcttgcc ttgagtgctt 9180
c 9181

<210> 15

<211> 29

<212> RNA

<213> Homo sapiens

<400> 15
ggcagaucug agccugggag cucucugcc 29

<210> 16

<211> 52

<212> RNA

<213> Homo sapiens

<400> 16
uuuuuuagg aagaucuggc cuuccuacaa ggaaaggcca ggaaauuuuuc uu 52

<210> 17

<211> 9413

<212> DNA

<213> Homo sapiens

<400> 17
ttggggcga cactccacca tagatcactc ccctgtgagg aactactgtc ttcacgcaga 60
aagcgtctag ccatggcggtt agtatgagtg ttgtcagcc tccaggaccc cccctcccg 120
gagagccata gtggctcg 180
gatccatca 240
tttttttgc 300
gttgggtcg 360
gttgggtcg 420
tttttttgc 480
gttgggtcg 540
tttttttgc 600

| | |
|--|------|
| tggggtgtggc aggatggctc ctgtcacccc gtggctctcg gcctagttgg ggcccccacag | 660 |
| accccccggcg taggtcgctgt aatttgggta aggtcatcgta tacccttaca tgcggcttcg | 720 |
| ccgacacctcat ggggtacatt ccgcttgcgt gcgcggccct agggggcgt gcccaggcccc | 780 |
| tggcacatgg tgtccgggtt ctggaggacg gcgtgaacta tgcaacaggg aatctgccc | 840 |
| gttgctcttt ctctatcttc ctcttagctt tgctgtcttg tttgaccatc ccagcttccg | 900 |
| cttacgaggt gcgcaacgtg tccggatata accatgtcac gaacgactgc tccaaactcaa | 960 |
| gtattgtgtta tgaggcagcg gacatgatca tgcacaccccc cgggtgcgtg ccctgcgtcc | 1020 |
| gggagagtaa tttctcccggt tgctgggtag cgctcaactcc cacgctcgcg gccaggaaca | 1080 |
| gcagcatccc caccacgaca atacgacgcc acgtcgattt gctcggttggg gcggctgctc | 1140 |
| tctgttccgc tatgtacgtt gggatctct gcgatccgt tttctcgtc tcccagctgt | 1200 |
| tcaccttctc acctcgccgg tatgagacgg tacaagattt caattgctca atctatcccg | 1260 |
| gccacgtatc aggtcaccgc atggcttggg atatgtatgtaa gaactggtca cctacaacgg | 1320 |
| ccctagtgggt atcgcagcta ctccggatcc cacaagccgt cgtggacatg gtggcgggggg | 1380 |
| cccaactgggg tgccttagcg ggccttgccc actattccat ggtggggaaac tgggctaagg | 1440 |
| tcttgattgt gatgtactc tttgctggcg ttgacgggca cacccacgtg acagggggaa | 1500 |
| gggttagcctc cagcacccag agcctcggt cctggctctc acaaggccca tctcagaaaa | 1560 |
| tccaaactcggt gaacaccaac ggcagctggc acatcaacag gaccgctctg aattgcaatg | 1620 |
| actccctcca aactgggttc attgctgcgc tttctacgc acacagggttc aacgcgtccg | 1680 |
| ggtgcccaga ggcgcgtggct agctgcccgc ccatcgatga gttcgcttag ggggtgggtc | 1740 |
| ccatcaactca tggatgtcct gagagctcg accagaggcc atattgttgg cactacgcgc | 1800 |
| ctcgaccgtg cgggatcggt cctgcgtcgc aggtgtgtgg tccagtgtat tgcttcactc | 1860 |
| cgagccctgt tggatgtgggg acgaccgatc gtttcggcgc tcctacgtat agctgggggg | 1920 |
| agaatgagac agacgtgtcg ctacttagca acacgcggcc gcctcaaggc aactggtttg | 1980 |
| ggtgacgtg gatgaacagc actgggttca ccaagacgtg cggggccct ccgtgcaaca | 2040 |
| tcgggggggt cggcaacaac accttggctc gccccacggta tgcttcggg aagcaccccg | 2100 |
| aggccactta cacaagggtt ggctcgggc cctgggttgc acccagggtgc atgggtgact | 2160 |
| accatacag gctctggcac taccctgca ctgttaactt taccgtcttt aaggtcaggaa | 2220 |
| tgtatgtggg gggcggtggag cacaggctca atgctgcgtg caattggact cgaggagac | 2280 |
| gctgtactt ggaggacagg gatagggtcg aactcagccc gctgctgtcg tctacaacag | 2340 |
| agtggcagat actgcctgt tccttcacca ccctaccggc cctgtccact ggcttgatcc | 2400 |
| atcttcacccg gaacatcggt gacgtgcaat acctgtacgg tataagggtcg gcagttgtct | 2460 |
| cctttgcaat caaatgggag tatacctgt tgctttcct tcttcggcg gacgcgcgcg | 2520 |

| | | | | | | |
|-------------|-------------|--------------|-------------|-------------|--------------|------|
| tctgtgcctg | cttgtggatg | atgtgtgtga | tagcccgagc | tgaggccacc | ttagagaacc | 2580 |
| tggtgtgcct | caatgcggcg | tctgtggccg | gagcgcatgg | ccttctctcc | ttcctcgtgt | 2640 |
| tcttctgcgc | cgcctggta | atcaaaggca | ggctggccc | tggggcggca | tatgtctct | 2700 |
| atggcgatg | gccgttgctc | ctgtcttgc | tggccttacc | accacgagct | tatgccatgg | 2760 |
| accgagagat | ggctgcatcg | tgcggaggcg | cggttttgt | agggtctggta | ctcttgacct | 2820 |
| tgtcaccata | ctataaggtg | ttcctcgcta | ggctcatatg | gtggttacaa | tattttatca | 2880 |
| ccagagccga | ggcgcacttg | caagtgtgg | tcccccctct | caatgttcgg | ggaggcccg | 2940 |
| atgccatcat | cctccttaca | tgcgcggtcc | atccagagct | aatctttgac | atcaccaaac | 3000 |
| tcctgctcgc | catactcggt | ccgctcatgg | tgctccaggc | tggcataact | agagtgcgt | 3060 |
| actttgtacg | cgttcagggg | ctcatccgt | catgcatgtt | agtgcggaag | gtcgctggag | 3120 |
| gccactatgt | ccaaatggcc | ttcatgaagc | tggccgcgt | gacaggtacg | tacgtatatg | 3180 |
| accatcttac | tccactgcgg | gattgggccc | acgcgggcct | acgagacett | gcggtgtggcag | 3240 |
| tagagccgt | cgtttctct | gacatggaga | ctaaactcat | cacctgggg | gcagacacccg | 3300 |
| cggcgtgtgg | ggacatcatc | tcgggtctac | cagtctccgc | ccgaaggggg | aaggagatac | 3360 |
| ttcttaggacc | ggccgatagt | tttggagagc | aggggtggcg | gctccttgcg | cctatcacgg | 3420 |
| cctattccca | acaaacgcgg | ggcctgcttg | gctgtatcat | caactgcctc | acaggtcggg | 3480 |
| acaagaacca | ggtcgatgg | gaggttcagg | tgctctccac | cgcaacgc当地 | tctttctgg | 3540 |
| cgacctgcgt | caatggcggt | tgttggaccg | tctaccatgg | tgcggctcg | aagaccctgg | 3600 |
| ccggcccgaa | gggtccaaatc | acccaaatgt | acaccaatgt | agaccaggac | ctcgtcggct | 3660 |
| ggccggcgcc | ccccggggcg | cgctccatga | cacccgtgcac | ctgcggcagc | tcggaccttt | 3720 |
| acttggtcac | gaggcatgt | gatgtcggtc | cggtgcgcgc | gcggggcgac | agcaggggga | 3780 |
| gcctgcttcc | ccccaggccc | atctcctacc | tgaagggctc | ctcgggtgga | ccactgcttt | 3840 |
| gcccttcggg | gcacgttgta | ggcatcttcc | gggctgctgt | gtgcacccgg | gggggtgcga | 3900 |
| aggcggtgga | cttcataaccc | gtttagtcta | tggaaactac | catgcggct | ccggcttca | 3960 |
| cagacaactc | atcccctccg | gccgtaccgc | aaacattcca | agtggcacat | ttacacgctc | 4020 |
| ccactggcag | cggcaagagc | accaaagtgc | cggctgcata | tgcagcccaa | gggtacaagg | 4080 |
| tgctcgctt | aaacccgtcc | gttgcgc当地 | cattgggctt | tggagcgtat | atgtccaagg | 4140 |
| cacatggcat | cgagccta | atcagaactg | ggtaaggac | catcaccacg | ggcggccca | 4200 |
| tcacgtactc | cacctattgc | aagttcccttgc | ccgacgggtgg | atgctccggg | ggcgctatg | 4260 |
| acatcataat | atgtgtatgaa | tgccactcaa | ctgactcgac | taccatcttgc | ggcatcgcc | 4320 |
| cagtcctgga | tcagggcagag | acggctggag | cgcgcgtcg | cgtgc当地 | accgcacgc | 4380 |
| ctccgggatc | gatcaccgtg | ccacacccca | acatcgagga | agtggccctg | tccaacactg | 4440 |

| | | | | | | |
|------------|-------------|-------------|--------------|-------------|-------------|------|
| gagagattcc | cttcttatggc | aaagccatcc | ccattgaggc | catcaagggg | ggaaggcattc | 4500 |
| tcatcttctg | ccattccaag | aagaagtgtg | acgagctcgc | cgcaaagctg | acaggcctcg | 4560 |
| gactcaatgc | tgttagcgtat | taccggggtc | tcgatgtgtc | cgtcataccg | actagcggag | 4620 |
| acgtcgttgt | cgtggcaaca | gacgctctaa | tgacggggtt | taccggcgcac | tttgactcag | 4680 |
| tgatcgactg | caacacatgt | gtcaccacaga | cagtgcattt | cagttggat | cccacccatca | 4740 |
| ccattgagac | gacaacgctg | ccccaaagacg | cggtgtcgcg | tgcgcagcgg | cgaggttagga | 4800 |
| ctggcagggg | caggagtggc | atctacaggt | tttgactcc | aggagaacgg | ccctcaggca | 4860 |
| tgttcgactc | ctcggtctcg | tgtgagtgtct | atgacgcagg | ctgcgccttgg | tatgagctca | 4920 |
| cgcggcgtga | gacctcggtt | aggttgcggg | cttacctaaa | tacaccaggg | ttgcccgtct | 4980 |
| gccaggacca | cctagagttc | tgggagagcg | tcttcacagg | cctcaccac | atagatgccc | 5040 |
| acttcttgc | ccagacccaaa | caggcaggag | acaacccccc | ctacctggta | gcataccaaag | 5100 |
| ccacagtgtg | cgcgcagggt | caggctccac | ctccatcg | ggaccaaattg | tggaagtgtc | 5160 |
| tcatacggct | aaagcccaca | ctgcatggc | caacgcctt | gctgtacagg | ctaggagccg | 5220 |
| ttcaaaatga | ggtcactctc | acacacccca | taaccaaata | catcatggca | tgcatgtcg | 5280 |
| ctgacctgga | ggtcgtca | agcacctggg | tgcttaggt | cggagtcctt | gcggctctgg | 5340 |
| ccgcgtactg | cctgacgaca | ggcagcgtgg | tcatgtggg | caggatcatc | ttgtccggga | 5400 |
| ggccagctgt | tattcccgac | aggaaagtcc | tctaccagga | gttcgatgag | atgaaagagt | 5460 |
| gtgcttcaca | cctcccttac | atcgagcaag | aatgcagct | cgcgcgacaa | ttcaaacaga | 5520 |
| aggcgctcgg | attgctgcaa | acagccacca | agcaagcgg | ggctgctgt | cccggtgtgg | 5580 |
| agtccaagtg | gcgagccctt | gaggcttct | ggggaaaca | catgtggaa | ttcatcagcg | 5640 |
| ggatacagta | cttggcaggc | ctatccactc | tgcctggaaa | ccccgcgata | gcatcattga | 5700 |
| tggctttac | agcctctatc | accagccgc | tcaccacccca | aaataccctc | ctgttaaca | 5760 |
| tcttgggggg | atgggtggct | gcccaactcg | ctccccccag | cgctgcttc | gtttcgtgg | 5820 |
| gcccggcat | tgccgggtgc | gccgtggca | gcataggct | cgggaaggta | cttggacac | 5880 |
| ttctggcggg | ctatggggcg | gggggtggct | gcccactcg | ggcctttaag | gtcatgagcg | 5940 |
| gcgcgcgtcc | ctccactgag | gatctggta | atttactccc | tgcctac | tctccgtgg | 6000 |
| ccctgggtgt | cggggcgtgt | tgcgcagca | tactgcgtcg | gcacgtggc | ccgggagagg | 6060 |
| gggcgttgca | gtggatgaac | cggctgatag | cggtcgcttc | cgggggtaac | cacgtctccc | 6120 |
| ccacgcacta | tgtgcccag | agcgacccg | cggcgcgtgt | tactcagatc | ctctccagcc | 6180 |
| ttaccatcac | tcaagggtgt | aagaggcttc | atcaggatggat | taatgaggac | tgctccacgc | 6240 |
| cttggccgg | ctcgtggcta | aaggatgtt | gggactggat | atgcacgggt | ttgagtgtact | 6300 |
| tcaagacttg | gctccagtc | aagtcctgc | cgcgttacc | gggactccct | ttcctgtcat | 6360 |

| | | | | | | |
|-------------|-------------|-------------|-------------|-------------|-------------|------|
| gccaacgcgg | gtacaaggga | gtctggcggg | gggatggcat | catgcaaacc | acctgcccatt | 6420 |
| gtggagcaca | gatcaccgga | catgtcaaaa | atggctccat | gaggattttt | ggccaaaaaa | 6480 |
| cctgcagcaa | cacgtggcat | ggaacattcc | ccatcaacgc | atacaccacg | ggccctgca | 6540 |
| cgccctcccc | agcgccgaac | tattccaggg | cgctgtggcg | ggtggctgct | gaggagtacg | 6600 |
| tggaggttac | gcgggtgggg | gatttccact | acgtgacggg | catgaccact | gacaacgtga | 6660 |
| aatgcccatt | ccaggttcca | gcccctgaat | ttttcacggg | ggtggatgga | gtacggttgc | 6720 |
| acaggtatgc | tccagtgtgc | aaacctctcc | tacgagagga | ggtcgatattc | caggtcgggc | 6780 |
| tcaaccagta | cctggtcggg | tcacagctcc | catgtgagcc | cgaaccggat | gtggcagtgc | 6840 |
| tcacttccat | gctcaccgac | ccctctcata | ttacagcaga | gacggccaag | cgtaggctgg | 6900 |
| ccagggggtc | tcccccctcc | ttggccagct | ttttagctag | ccagttgtct | gccccttctt | 6960 |
| tgaaggcgac | atgtactacc | catcatgact | ccccggacgc | tgacctcatc | gaggccaacc | 7020 |
| tcctgtggcg | gcaggagatg | ggcgggaaca | tcaccgtgt | ggagtcagaa | aataagggtgg | 7080 |
| taatccttgg | ctcttcgtat | ccgattcggg | cggggaggg | ttagggggaa | atatccgtcc | 7140 |
| cggcggagat | cctgcggaaaa | cccgaggaaat | tcccccagc | gttggccata | tggcacgccc | 7200 |
| cggattacaa | ccctccactg | ctagacttcc | ggaaggaccc | ggactacgtc | ccccgggtgg | 7260 |
| tacacgggtg | ccctttgcca | tctaccaagg | ccccccaaat | accacctcca | cgaggaaaga | 7320 |
| ggacgggtgt | cctgacagag | tccaccgtgt | tttctgcctt | ggcggagctc | gctactaaga | 7380 |
| cctttggcag | ctccgggtcg | tcggccgttg | acagggcac | ggcgaactggc | cctcccgatc | 7440 |
| aggcctccga | cgacggcgac | aaaggatccg | acgttgagtc | gtactcctcc | atgccccccc | 7500 |
| tcgagggaga | gccaggggac | cccgacctca | gacgacgggtc | ttggtctacc | gtgagcgggg | 7560 |
| aagctggtga | ggacgtcgac | tgctgctcaa | tgtcctatac | atggacaggt | gccttgcata | 7620 |
| cgcctatgcgc | tgccggaggag | agcaagttgc | ccatcaatcc | gttggcaac | tctttgctgc | 7680 |
| gtcaccacag | tatggtctac | tccacaacat | ctcgacgcgc | aagtctgcgg | cagaagaagg | 7740 |
| tcaccttga | cagactgcaa | gtcctggacg | accactaccg | ggacgtgctc | aaggagatga | 7800 |
| aggcgaaggc | gtccacagtt | aaggcttaggc | ttcttatctat | agaggaggcc | tgcaaactga | 7860 |
| cgcggccaca | ttcggccaaa | tccaaatttg | gctacggggc | gaaggacgtc | cgaggcctat | 7920 |
| ccagcagggc | cgtcaaccac | atccgcctcg | tgtggagga | tttgcgtggaa | gacactgaaa | 7980 |
| caccaattga | taccaccatc | atggcaaaaa | atgaggtttt | ctgcgtccaa | ccagagaaag | 8040 |
| gaggccgcaa | gccagctcgc | tttatctat | tccagacct | gggggtacgt | gtatgcgaga | 8100 |
| agatggccct | ttacgacgtg | gtctccaccc | ttcctcaggc | cgtgtatgggc | ccctcatacg | 8160 |
| gattccagta | ctctcctggg | cagcgggtcg | agttccgtgt | gaatacctgg | aaatcaaaga | 8220 |
| aatgccttat | gggcttctca | tatgacaccc | gctgtttga | ctcaacggtc | actgagaatg | 8280 |

<210> 18

<211> 328

<212> RNA

<213> *Homo sapiens*

<400> 18

uuggggcgca cacuccacca uagaucacuc cccugugagg aacuacuguc uucacgcaga 60
 aagcgucuag ccauggcguu aguaugagug uugugcagcc uccaggaccc ccccuccccgg 120
 gagagccaua guggucugcg gaaccgguga guacaccgga auugccagga cgaccggguc 180
 cuuuucuugga ucaacccgcu caauggcugg agauuugggc gugccccccgc gagacugcua 240
 gccgagugu guggggucgc gaaaggccuu gugguacugc cugauagggu gcuugcgaguu 300
 gccccgggaaq gucucguaga ccquqcau 328

<210> 19

<211> 14

<212> RNA

<213> Homo sapiens

<400> 19

auuugggcgu gcc

14

<210> 20

<211> 27

<212> RNA

<213> Homo sapiens

<400> 20

gccgaguagu guugggucgc gaaaggc

27

<210> 21

<211> 340

<212> DNA

<213> Homo sapiens

<400> 21

atgggcggag ggaagctcat cagtgggcc acgagctgag tgcgtcctgt cactccactc 60

ccatgtccct tgggaaggc tgagactagg gccagaggcg gccctaacag ggctctccct 120

gagcttcagg gaggtgagtt cccagagaac ggggctccgc gcgaggtcag actggcagg 180

agatgccgtg gaccccgccc ttccgggagg ggcccgccgg atgcctcctt tgccggagct 240

tggAACAGAC tcacggccag cgaagtgagtt tcaatggctg aggtgaggtt ccccgccagg 300

gacctataa cccaaattcag accactctcc tccggccatt 340

<210> 22

<211> 349

<212> DNA

<213> Homo sapiens

<400> 22

| | | | | | | |
|------------|------------|-------------|-------------|------------|------------|-----|
| gaggaaagtc | cgggctcaca | cagtctgaga | tgattgttagt | gttcgtgctt | gatgaaacaa | 60 |
| taaatcaagg | cattaatttg | acggcaatga | aatatcctaa | gtctttcgat | atggatagag | 120 |
| taatttggaa | gtgccacagt | gacgttagctt | ttatagaaat | ataaaaggtg | gaacgcggta | 180 |
| aaccctcga | gtgagcaatc | caaatttggt | aggagcactt | gtttaacgga | attcaacgta | 240 |
| taaacgagac | acacttcgct | aaatgaagtg | gtgtagacag | atggttatac | cctgagtacc | 300 |
| agtgtgacta | gtgcacgtga | tgagtacgat | ggaacagaac | gcggcttat | | 349 |

<210> 23

<211> 377

<212> DNA

<213> Homo sapiens

| | | | | | | |
|------------|------------|-------------|------------|------------|------------|-----|
| <400> 23 | | | | | | |
| gaagctgacc | agacagtcgc | cgttcgtcg | tcgtcctctt | cgggggagac | ggcgaggagg | 60 |
| gaggaaagtc | cgggtccat | aggcagggt | gccaggtaac | gcctgggggg | gaaaccacg | 120 |
| accagtgcaa | cagagagcaa | accgcccgtat | gcccgccaa | gcgggatcag | gtaagggtga | 180 |
| aagggtgcgg | taagagcgca | ccgcgcggct | ggtaacagtc | cgtggcacgg | taaactccac | 240 |
| ccggagcaag | gccaaatagg | ggttcataag | gtacggcccg | tactgaaccc | ggtaggctg | 300 |
| cttgagccag | tgagcgattt | ctggcctaga | tgaatgactg | tccacgacag | aaccggctt | 360 |
| atcggtcagt | ttcacct | | | | | 377 |

<210> 24

<211> 38110

<212> DNA

<213> Homo sapiens

| | | | | | | |
|------------|------------|------------|------------|-------------|------------|-----|
| <400> 24 | | | | | | |
| ccaccggta | cgatcttgcc | gaccatggcc | ccacaatagg | gccggggaga | cccgccgtca | 60 |
| gtggtggcg | gcacggtcag | taacgtctgc | gcaacacggg | gttgaactgac | ggcaataatc | 120 |
| ggctccatag | cgtcgccgc | ggatacagta | aaggagcatt | ctgtgacgga | aaagacgccc | 180 |
| gacgacgtct | tcaaacttgc | caaggacgag | aaggtcgaat | atgtcgacgt | ccggttctgt | 240 |
| gacctgcctg | gcatcatgca | gcacttcacg | attccggctt | cggccttga | caagagcgtg | 300 |
| tttgacgacg | gcttggcctt | tgacggctcg | tcgattcgcg | ggttccagtc | gatccacgaa | 360 |
| tccgacatgt | tgcttcttcc | cgatcccgag | acggcgcgca | tcgaccgggtt | ccgcgcggcc | 420 |

| | | | | | | |
|-----------------------|-----------------------|------------|------------|------------|------------|------|
| aagacgctga atatcaactt | ctttgtcac gaccggttca | ccctggagcc | gtactcccg | 480 | | |
| gaccggcgca acatcgcccc | caaggccgag aactacctga | tcagcactgg | catcgccgac | 540 | | |
| accgcatact | tccggcgccga | ggccgaggtc | tacatttcg | attcggtgag | 600 | |
| cgcgcacaacg | gctccttcta | cgaggtggac | gccatctcg | ggtgtggaa | caccggcg | 660 |
| gcgaccgagg | ccgacggcag | tcccaaccgg | ggctacaagg | tccgccacaa | gggcgggtat | 720 |
| ttcccagtgg | cccccaacga | ccaatacg | tcacgtgc | acaagatgt | gaccaac | 780 |
| atcaactccg | gcttcatcct | ggagaagggc | caccacgagg | tgggcagcgg | cgacaggcc | 840 |
| gagatcaact | accagtcaa | ttcgctgct | cacgcccgg | acgacatgc | gttgtacaag | 900 |
| tacatcatca | agaacaccgc | ctggcagaac | ggcaaaacgg | tcacgttcat | gcccaagccg | 960 |
| ctgttcggcg | acaacgggtc | cgcatgc | tgtcatcagt | cgctgtggaa | ggacggggcc | 1020 |
| ccgctgatgt | acgacgagac | gggttatgcc | ggtctgtcg | acacggcccg | tcattacatc | 1080 |
| ggcggcctgt | tacaccacgc | gccgtcgct | ctggccttca | ccaacccgac | ggtgaactcc | 1140 |
| tacaagccgc | tggttcccg | ttacgaggcc | ccgatcaacc | tggctatag | ccagcgcaac | 1200 |
| cggtcggcat | gcgtgcgc | cccgatcacc | ggcagcaacc | cgaaggccaa | gcccgtggag | 1260 |
| ttccgaagcc | ccgactcg | ggcaacccg | tatctggcgt | tctcgccat | gctgatggca | 1320 |
| ggcctggacg | gtatcaagaa | caagatcgag | ccgcaggcgc | ccgtcgacaa | ggatctctac | 1380 |
| gagctgccgc | cggaaagaggc | cgcgagtatc | ccgcagactc | cgacccagct | gtcagatgt | 1440 |
| atcgaccgtc | tcgaggccga | ccacgaatac | ctcaccgaag | gaggggtgtt | cacaacgac | 1500 |
| ctgatcgaga | cgtggatcag | tttcaagcgc | gaaaacgaga | tcgagccgt | caacatccgg | 1560 |
| ccgcattccct | acgaattcgc | gctgtactac | gacgtttaag | gactcttcgc | agtccgggt | 1620 |
| tagagggagc | ggcgtgtcg | tgccagggcg | ggcgtcgagg | ttttcgatg | ggtgacgggt | 1680 |
| gccggcaacg | gcgcgcccac | caccgtcg | aagagccgt | ttaagaacgt | tcaaggacgt | 1740 |
| ttcagccgg | tgccacaacc | cgcttggcaa | tcatctcccg | accggcagc | gggttgtctt | 1800 |
| tcacatgcgc | cggaaactcaa | gccacgtcg | cgcccaggcg | tgcgtcg | gccgggtcag | 1860 |
| gttaagtgtc | ggggattcgt | cgtgcggcg | ggcgtccacg | ctgaccaacg | gggcagtcaa | 1920 |
| ctcccaaca | cttgcgcac | taccgcctt | gcccggcg | tcacccgtag | gtagttgtcc | 1980 |
| aggaattccc | caccgtcg | gttgcgcac | ccggccgcga | ccgcgaccgc | attgagctgg | 2040 |
| cggccgggtc | ccggcagctg | gtcggtggc | ttgcccgc | ccaacaccag | cgcgttgcgg | 2100 |
| gcccgggtgg | cggtcagcca | ggcctgacgg | agcagctcca | cgtcggctgc | gggaaccaga | 2160 |
| tcggcggccg | cgatgacatc | cagggattgc | agcgtcgagg | tgttgtgc | ggcgggaacc | 2220 |
| tggtgccat | gctgttagctg | cagcaactgc | acggtccatt | cgatgtcg | cagtccgc | 2280 |
| cggcccaagtt | tggtgtgtgt | gttgggtcg | gcaccgcg | gcaaccgctc | ggactcgata | 2340 |

| | | | | | | | | |
|-------------|-------------|-------------|-------------|------------|------------|------------|------------|------|
| cgggccttga | tgccgcgaat | ctcgccacc | gagtca | gcgg | acacaccgtc | ggcg | gata | 2400 |
| cgcgtttgt | cgaccatccg | taggaatcgc | tgacccaact | cgccatcgcc | ggcaaccgcg | | | 2460 |
| tgtgcgcgt | gcagggcctg | gatctccat | ggctgtccc | actgctcgta | gtatgcggcg | | | 2520 |
| taggaccca | gggtgcggac | cagcggaccg | ttgcggccct | cggtcgcaa | attggcgtcg | | | 2580 |
| agotccagcg | gcggatcgac | gctgggtgtc | cccagcagcg | cccgaacccg | ctcggcgatc | | | 2640 |
| gatgtcgacc | atttcaccgc | ccgtgcac | tcgacgcccgg | tggccggctc | acagacgaac | | | 2700 |
| atcacgtcgg | catccgaccc | gtagccaaac | tcggcaccac | ccagccgacc | catgcccgt | | | 2760 |
| accgcgatgg | ccgcgggggc | gcgatcg | tcgggaaggc | tggccggat | catgacgtcc | | | 2820 |
| agcgcggcct | gcagcaccgc | cacccacacc | gacgtcaac | ccggcacac | ctcggtgacc | | | 2880 |
| tcgagcaggc | cgagcaggc | cgccgaaccg | atgcggcca | gctctcgac | acgcagcgt | | | 2940 |
| cgcgccgg | cgatggcccg | ctccgggtcg | ggtagcggc | tcgcccaggc | gatcagcgcc | | | 3000 |
| cgagccacgg | cgccgggctc | ggtctcgagc | agcttccggc | ccgcaggccc | gtctctgtac | | | 3060 |
| tgctggatga | cccgccggcgc | g | gcgcataac | agatccggca | catacgccga | gttacccaag | | 3120 |
| acatgcatga | gccgcttggc | caccgcggc | ttgtcccgca | gcgtggccag | gtaccagctt | | | 3180 |
| tcggtgccca | g | gcgcctca | gagccgcccgg | taggcagca | gtccgcgtc | ggatcgggg | | 3240 |
| gcatacgaca | tccagtccag | cagcctggc | agcagcaccg | actgcacccg | tccgcggcgg | | | 3300 |
| ccgctttgat | tgaccaacgc | cgacatgtgt | ttcaacgcgg | tctgcgtcc | ctcgtagccc | | | 3360 |
| agcgcggcca | gccggcgccc | cgccgcctcc | aacgtcatgc | cgtggcgtat | ctccaacccg | | | 3420 |
| gtcgggcccga | tcgattccag | cagcggttga | tagaagagtt | tggtgtttaa | tttcgacacc | | | 3480 |
| cgcacgttct | gcttcttgag | ttcctcccg | agcaccccg | ccgcac | tcggccatcg | | | 3540 |
| ggccggatgt | ggccgcgcg | cgccagccag | cgca | ctcgcttc | ggatcggga | | | 3600 |
| agcaggtggg | tgcgtttag | ccgctgcaac | tgcagtgcgt | gctcgagcag | cctgaggaac | | | 3660 |
| tcatacgacg | cggtcatgtt | cgccgcgtcc | tcacgcccga | tgtagccg | ttcgcccaac | | | 3720 |
| gccgccaatg | cgtccaccgt | ggacgcccacc | cgtaacgact | cgtcgctacg | ggcatgaacc | | | 3780 |
| agctgcagta | gctgtacggc | gaactccacg | tcgcgcata | cgccgctg | gagttgagc | | | 3840 |
| tcgcggccgc | ggacatcg | gggcacc | tgctccaccc | gccgccc | gacatcgacc | | | 3900 |
| tcgaccacaa | agtcttcgc | ctcg | cagg | cgcc | acacca | tcggcatcaa | ggcggtcagg | 3960 |
| taacgctcgc | caagttccgc | gtcg | ccaa | actggccgt | ctt | cagcaa | cgctgaaac | 4020 |
| tcccaggtct | tggcc | ccagcg | ctgg | tagtag | gcgatgtcg | actcgagcgt | acggacc | 4080 |
| tccccgttgc | gcc | cctccgg | acgc | agg | cg | gtccac | cgaaaaaggc | 4140 |
| acccgcac | tctcg | cttggc | cac | gcgc | gcgc | ttgcgcgg | cgagcgc | 4200 |
| atgacatcg | cg | tcgt | cgac | gtat | tcgc | gcac | cgacttgc | 4260 |

| | | | | | | |
|-------------|-------------|------------|------------|-------------|-------------|------|
| accggccaggc | gcgggtggcgg | gtgctcgccg | cacacgctcg | cctcgccac | gcgcagcgcc | 4320 |
| gccgccagag | cggcgccgc | ggcgccgc | aggcggtcg | ccaccacgt | aatggcagc | 4380 |
| accgggtcg | cctcgaccgt | cgcggccagg | tcgagagcg | ccagcattag | cacgtagtcg | 4440 |
| cggtaactgg | ttcgcaatcg | gtgcacgagc | gagcccgca | taccctccga | ttcctcgacg | 4500 |
| cactcgacga | acgaccgctg | cagctggtca | tgggacggca | gtgtgacctt | gccccgcagc | 4560 |
| aatttccagg | actgcggatg | ggcgaccagg | tgatcgccc | acgcccagcga | cgagcccagc | 4620 |
| accgagaaca | gccgcggcg | cagactgcgt | tcgcgcagca | gagccgcgtt | gagctcggtcc | 4680 |
| catccggtgt | ctggattctc | cgacagccgg | atcaaggcgc | gcagcgcggc | atcggcgtcc | 4740 |
| ggagcgcgtg | acagcgacca | cagcaggctg | acgtgcgcct | gatcctcg | ccgatcccac | 4800 |
| cccaagctgag | ccagacgctc | accagcaggg | gggtcaacta | atccgagccg | gccaacgcgt | 4860 |
| ggcaacttcg | gccgctgcgt | ggcgagttt | gtcacgacca | cgacggtagc | gcaaagcg | 4920 |
| tcggcgtcgg | atcaaccgg | agatctggc | tacagcgaca | ggttaggtgc | cagctcgat | 4980 |
| ggcgtgacgt | ggctgcggta | gttcgcccac | tccgtgcgt | tgttgcgcaa | aaaaaagtca | 5040 |
| aaaacgtgct | cccccaaggc | ctccgcgacg | agttcgagg | cctccatggc | gcgcagcgca | 5100 |
| ctatccaaac | tggacggcaa | ttctcggtac | cccatcgctc | ggcgttcctc | gggtgtgagg | 5160 |
| tcccatacgt | tgtcctcg | ctgcgggccc | agcacgtaac | ccttcctac | accccgcaat | 5220 |
| cccgccgcca | gcagcacggc | aatgtcaga | taggattgc | acgcccatac | agggctgcgt | 5280 |
| acttcgaccc | gccgcgacga | gttcttg | ggcgtgtaca | tcggcaccc | cactagggcg | 5340 |
| gatcggttgg | cggcccccca | cgacgcggc | gtggcgctt | cggccctg | caccagccgc | 5400 |
| ttgtaaagat | tgacccactg | atttgtgacc | gacgtgatct | cgcaagcg | ctccaggatc | 5460 |
| ccggcgatga | acgatttacc | cacttccgac | agctgcagc | gatcatc | gctgtgaa | 5520 |
| gcgttgcacat | caccctcgaa | caggctcatg | tgggtgtca | tcgcccagcc | cggtgtctgg | 5580 |
| ccgaatggct | tggcatgaa | cgacgcccgg | gcccctt | ccagcgc | ttcttgat | 5640 |
| acgtacgg | aggcatcac | gttgtcagcc | atcgacagag | cgtcgcaaa | ccgcagg | 5700 |
| atctcctgct | ggccgggtgc | gccttcgt | tggctgact | ccaccgagat | gcccata | 5760 |
| tccagggcat | cgatcg | cgccgc | aaag | ttcaagg | agtcgtgcac | 5820 |
| aaatagccgg | cgttgc | cgggacgggc | accgaccgt | cctcggt | ggcttg | 5880 |
| aggaagaact | cgttgcgg | atgcacgt | caggagaagc | cgagttcg | ggcattcg | 5940 |
| agctgcc | gcaacacgt | ccgcgggt | gccacgac | gcgagcc | cgcatgg | 6000 |
| atgtcgaaa | acatccgc | tgagtgg | tggccgaa | tgtggcc | ggcagc | 6060 |
| tggaaagg | acgggtccgg | gtgc | ccacc | gtatcg | ccgagaccc | 6120 |
| tcgatcgagg | atccgtcgaa | gccgatgc | ctctcg | aagg | cgcaaagccc | 6180 |

| | |
|--|------|
| gcgatggcga ccgacttgag gaaaccgagc acgtctgtga accacagccg gacgaagcgg | 6240 |
| atgtcgctt cttccagggt acgaagaacg aattccttct gtcggtccat acctcgaaca | 6300 |
| gtatgcactg tctgttaaaa ccgtgttacc gatgcccggc cagaagcgtt gcggggcggc | 6360 |
| ccgcaagggg agtgcgcggt gagttcaggg cgccgaccgc agactcgctg gcggcaaggt | 6420 |
| cccgtcgaga aaatagtgc a tcaccgcaga gtccacacac tggttgcat cgaacaccgc | 6480 |
| agtgtgttgg gtgccgtcga aggtgatcag cggtgcgccc agctggcggg ccaggctac | 6540 |
| cccgactga tacggagtgg ccgggtcgta ggtggtggac accacgacga cttgccagc | 6600 |
| cccgccggc gccgcgggt gcggcgtcga cgttgccggc accggccaca gcgcgcacag | 6660 |
| atcgcggggg gcgatccgg tgaactgccc gtagctaagg aacggggcga cctgacggat | 6720 |
| ccgttggtcg gcggccaccc aggccgtgg atcggccgt gtgggcgtat cgacgcaccc | 6780 |
| gaccgcgttg aacgcgtctt ggtcggtct gtagtgcggc tctgcattcc ggcgtcata | 6840 |
| gtcgtcggca agcaccagca agtcgcggc gtcgctgcg cgctgcagcc ccagcagacc | 6900 |
| actggtcagg tacttccagc gctgagggt gtacagcgcg ttgatggtgc cctgcgtcgc | 6960 |
| gtcggcgttag ctcaggccac gtggatccga cgtcttaccc ggcttctgca ccagcgggtc | 7020 |
| aaccagggcg tggtagcggt tgacccactg ggccgagtcg gtgcccagag ggcaggccgg | 7080 |
| cgagcggcgg cagtcggcgg cgtagtcatt gaaagcggtc taaaatcccg ccatttggct | 7140 |
| gatgctttcc tggatgggc taacggctgg atcgatagcg cgtcgagga ccatcgcccg | 7200 |
| cacatgagta cggaaaccgtt ccaggtaaagc ggtgcccac tcggtgcgt agctgtatcc | 7260 |
| gaggttagtt atctgatcgt cacctaacgc ttggcgaacc atgtccatgt cccgtgcgac | 7320 |
| ggacgcggta cccatattgg ccaagaagct gaagcccatc cggtaacac agtcctggc | 7380 |
| caactgccgg tagacctgtt cgacgtgggt gacacggcc ggactgttagt cggccatcg | 7440 |
| atcgcgccgg tacgcgtcga actcggcgtc ggtgcgacac cgaacgcag ggtcgagtg | 7500 |
| gccgaccctt ctgggtcga agcccaccag gtcgaagtgg cggagaatgt cgggtcgcc | 7560 |
| gatcgcggtt gccatagcgg cgaccatgtc gaccgcgcac gccccgggtc cccaggatt | 7620 |
| gaccagcagt gctccgaatc gctgtccgt cgcggggacg cggatcaccc ccaacttcgc | 7680 |
| ttgtgtccca cgggttgggt cgtagtcgac gggacggac accgtcgccg acgtgcagt | 7740 |
| gcgaatttcg ctgggtcgg cgatgaactc cggcggacgt ttccaaactct gttgcggcgc | 7800 |
| cacgaccggc gcacccgggg tttggccggc gccgggttct tcagtcgcgc cggcaacgg | 7860 |
| gggcgtgtct aggggcagtc cggcggacgt caacccgaag gacagcagcg cggagctaa | 7920 |
| cggctcgcc cgccacatgg cggccatcg ttcaccggcg aatacctgtg acggcgcgaa | 7980 |
| atgatcacac cttcggttct tcgccccgt agcacttggc gccgctggc ggcgtggtgc | 8040 |
| cgccgattaa atacgcgtc acgtactcgtaatgcagtc gtcgcccgtt aataccaccg | 8100 |

| | |
|--|-------|
| tgtgctgggt tccgtcgaag gtcagcaacg aaccgcgaag ctggttcgcc aggtcgaccc | 8160 |
| cggcattgtta cggcgtcgccc gggtcatggg tggtgatac caccaccgtc ggcactaggc | 8220 |
| cgggcgcgca gacggcatgg ggctgacttg tgggtggcac cggccagaac ggcaggtgc | 8280 |
| ccagcggcgc atcaccggtg aacttcccgta agctcatgaa cggtgcgatc tcccgccgc | 8340 |
| ggcggctttc gtcgatgacc ttgtcgcgat cggtaaccgg gggctgatcg acgcaattga | 8400 |
| tgcgcaccccg cgcgtcacccg gaattgttgt agcggccgtg cgagtcccgaa cgcatgtaca | 8460 |
| tgtcggccag agccagcagg gtgtctccgc gattgtcgac cagctccgac agcccgctgg | 8520 |
| tcaagtgttg ccacagattc ggtgagtaca ggcgcataat ggtgcccacg atggcgctgc | 8580 |
| tataactcag cccgcgcgga tccttcgtgc ggcgcggcct gctgatcctc gggttgtccg | 8640 |
| ggtcgaccaa cggatcgacc aggctgttgtt agacctcgac ggctttggcc gggtcggcgc | 8700 |
| ccagcgggca gcccgcgttc ttggcgcaagt cggcggcata gttgttgaac gcgtcctgga | 8760 |
| agcccttggc ctggcgccgc tccgcctcga tgggatcgac attgggttcg acggcaccgt | 8820 |
| cgagaatcat tgcggcacc cgtcgccgaa attcctcgac atacgcggag ccgatccggg | 8880 |
| tgcgtacga gtagcccgagg taggtcagct tgcgtcgcc caacgcgcg cgaatggcat | 8940 |
| ccaggtccctt ggcgacgttg accgtcccgaa catgggccag aaagttcttgc cccatcttgc | 9000 |
| ccacacagcg accgacgaat tgcttggctc cggtctcgat gtgcgcacca ccctccggc | 9060 |
| tgttagtcaac ctgcggctcg gcccgcagcc ggtcggtgtc ggcacatcgag ttgcaccaga | 9120 |
| tcgcccggccg ggacgacgc accccgcggg ggtcgaaccc aaccagggtcg aacctttcg | 9180 |
| gcacccgctt cggcaatgtc tggaaagacgc ccaaggcggc ctcgataaccg gattgcggg | 9240 |
| gtccaccggg atttatgacc agcgaaccga tcttgcgtcc cgtcgccgaa aagcgaatca | 9300 |
| gcccgcgcgc cggccacgtca ccatcggggc ggtcgtagtc gaccggatc gcgagcttgc | 9360 |
| cgcataacgc gcccgcgggg atctttactt gcggtttga cggccgcac ggtgtccact | 9420 |
| ccaccggctg gcccagcttc ggctccgcca tacgagcgcg tccccgacc acgcggatgc | 9480 |
| agcccacaag aaccaacgc accggcggcga ggcgcggccca gatcaacacg atgcgcgcga | 9540 |
| tcttgcgtcg gcgagacacgc ctcatgcggca caatgcgtcc agagcagacc cgagatcctg | 9600 |
| gccagcggcc accgtcgccg gactaaccgg cggctgcccag cagtcctgccc atgcggatgc | 9660 |
| cgaaactcgt cggccatccc ccatacgcc ggtaaacagat cccggcaaga caccgaccgg | 9720 |
| tcgaccggat cgggcacggg cgcgtcgcc tcggcggtgc acaactgcga catcagggttgc | 9780 |
| gcgctggcac cccgtccacg cggcatggt gaccccttgc catcgccca gggcgatccc | 9840 |
| cgatgcgcgc cacccttcg acgaaccat ctcccacggc ggtcgccggc agcgacgcga | 9900 |
| tgtggccgca gatctcccgag agttcgcccc gccccccgg cgacggcaac ccgatgcgttgc | 9960 |
| gcaagtgcacg atcgatgtga ggttcaaggt tcagcgcact gctggcaagc tttttccgaa | 10020 |

accgcggcct cgccttgate tggagtca aacgcgtc aacgcgtca aaggcgtaac 10080
ccatgctcga gcaaacatgc atgggctgag tggacgttc cagacacagc aactggcgtc 10140
caggccactg agccgctgca tgcgcgtatgg tatgccatgg gggcccccgg gcgcgtctga 10200
gggaaagaag tggcagactg tcagggtccg acgaacccgg ggaccctaac gggccacgag 10260
gatcgacccg accaccatta gggacagtga tgtctgagca gactatctat ggggccaata 10320
ccccggagg ctccggccg cgaccaaga tccgcaccca ccacctacag agatggaagg 10380
ccgacggcca caagtggcc atgctgacgg cctacgacta ttcgacggcc cgatcttcg 10440
acgaggccgg catccgggtg ctgctggcgt gtgattcggc ggcaacgtc gtgtacggct 10500
acgacaccac cgtgcccgtc tccatcgacg agctgatccc gctggtccgt ggcgtggc 10560
ggggtgcccc gcacgcactg gtcgtcgccg acctgcccgt cggcagctac gaggcggggc 10620
ccaccggccgc gttggccgccc gccacccggt tcctcaagga cggcggcgca catcggtca 10680
agctcgaggg cggtgagcgg gtggccgagc aaatcgctg tctgaccgcg gcggcatcc 10740
cggtgatggc acacatcgcc ttcacccgc aaagcgtcaa caccttggc ggcttccggg 10800
tgcaggcccg cggcgacgccc gccgaacaaa ccatcgccga cgatcgcc gtcgcccgaag 10860
ccggagcggt tgccgtcgatgg atggatgg tgccgcga gttggccacc cagatcaccg 10920
gcaagcttac cattccgacg gtcggatcg gcgtggcccaactgcgac ggccaggtcc 10980
tggatggca ggacatggcc ggttcagcg gcgcacagac cggccgttc gtcaaacgg 11040
atgcccgtgt cggtggtgaa ctacggcgatgg ctgcaatgca atacgcccggg 11100
gcgggttatt ccccgctgac gaacacagtt tctgaccaag ccgaatcagc cggatcgccg 11160
ggcattgcgg tggcgccctg gatgccgtcg acggcgatt gccggcgccg acgcggccagc 11220
gggacccatc ggcgtcgctg tcgcccgtt agccgggggt gagcccagac attcgatgt 11280
cccaacacca tccgcccacag cccaaattgtat gtggcactct atgcgtgcct atccccgacc 11340
aaccaccacc gcggcgacgc atcatgaccg gaggcgaaga tgccagttaga ggcccccaga 11400
ccagcgcgccc atctggaggt cgagcgcaag ttgcacgtga tcgagtcgac ggtgtcgccg 11460
tcgttcgagg gcatcgccgc ggtggttcgc gtcgagcgt cgccgaccca gcagctcgac 11520
gcgggttact tcgacacacc gtcgcacgac ctggcgccga accagatcac cttgcggcg 11580
cgcacccggcg gcggcgacgc cggctggcat ctgaagctgc cggccggacc cgacaagcgc 11640
accgagatgc gagcaccgatgtat gcgcacgtg tgccggccga gttgttggat 11700
gtgggtctgg cgatcgatcccg cgaccaggcg gttcagccgg tcgcccggat cagcactcac 11760
cgcgaaagcc agatcctgtat cggcgccggg ggcgacgcgc tggcgaaatt ctgcaacgac 11820
gacgtcaccg catggtcggc cggggcattc cacggcgatgtat gtgcagcgaa caacggccct 11880
gccgaacacgc agtggcgccga atggaaactg gaactggatca ccacggatgg gaccggccat 11940

accaagctac tggaccggct agccaaccgg ctgctcgatg ccgggtgccgc acctgccggc 12000
cacggctcca aactggcgcg ggtgctcggt gcgacctctc ccgggtgagct gccaacggc 12060
ccgcagccgc cggcggatcc agtacaccgc gcgggtgtccg agcaagtgcg gcagctgctg 12120
ctgtgggatc gggccgtgcg ggccgacgcc tatgacgccc tgcaccagat gcgagtgacg 12180
acccgcaaga tccgcagctt gctgacggat tcccaggagt cggttggcct gaaggaaagt 12240
gcgtgggtca tcgatgaact gcgtgagctg gccgatgtcc tgggcgttagc ccgggacgcc 12300
gaggtactcg gtgaccgcta ccagcgcgaa ctggacgcgc tggcgcggaa gctggtacgc 12360
ggccgggtgc gcgagcgcct ggtagacggg gcgcggcgcc gataccagac cgggctgcgg 12420
cgatcaactga tcgcattgcg gtcgcagcgg tacttccgtc tgctcgacgc tctagacgc 12480
cttgtgtccg aacgcgccc tgcacttct ggggaggaat cggcaccggt aaccatcgat 12540
gcggcctacc ggcgagtccg caaagccgca aaagccgcaa agaccgcgg cgaccaggcg 12600
ggcgaccacc accgcgacga ggcattgcac ctgatccgca agcgcgcgaa gcgattacgc 12660
tacaccgcgg cggctactgg ggccgacaat gtgtcacaag aagccaaggt catccagacg 12720
ttgcttaggcg atcatcaaga cagcgtggc agccggaaac atctgatcca gcagggcata 12780
gccgcgaaca ccgcggcgaa ggacacccctc acctacggtc tgctctacca acaggaagcc 12840
gacttggccg agcgcgtccg ggagcagctt gaagccgcgc tgcgcaact cgacaaggcg 12900
gtccgc当地 caacgggattt agccgc当地 gggcggacga gttggcctgt aagccggatt 12960
ctgttccgc当地 ccgc当地 acaggc当地 caagctaact gcggcacggc ggccgaccatc catctggaca 13020
caccgttacc gggc当地 cccgc当地 agccgc当地 ccgc当地 ggccgagcaa ccctcaagcg 13080
cctgc当地 ccgc当地 cgc当地 acttccggtt gtgc当地 cttgc当地 cttccgggtt ggcc当地 ccttgc当地 13140
gccacccccc tcacccggaa tgctggc当地 ctcttaccgc accgtttcac cttgccacc 13200
acgaggatgg cggctgttt tctgtggc当地 tttccgc当地 gtc当地 ctc当地 attgccgtta 13260
gcaatcaccc tgctctgtga agtccggact ttcctcgact cgacgctgaa cctcgtaat 13320
ccacacaaggc cctacgc当地 ccgc当地 ccgc当地 ccagccact catccgc当地 gaccacgcta 13380
ccccgctggg cggc当地 cggc当地 ggc当地 cggc当地 acacggctag tcggacagcc 13440
gatccggccgg gcagtc当地 tcgtggactg gtgacacgggt gggacaaacg cgtc当地 actcc 13500
ggcgactggg acgccatcgc tgccgaggc当地 agc当地 agtgc当地 gttggc当地 cacttc当地 cgg 13560
ctgatcaccc ccggc当地 gaggc当地 cggccggctg cgc当地 agtgc当地 acgccc当地 acgca 13620
cgctcgacgg tcgatatggc atccaaggc当地 tacggc当地 cccggctg ggc当地 cgtatcg atattccat 13680
gccccctatc ccgagtgatc gagcgtctca agcaggc当地 gtc当地 cccaaa ctgctgccg 13740
tagcgc当地 cccggccaa ctggtggccaa actggccggcc gggaggc当地 ctggccagac agccttgatg 13800
actggc当地 gagctgtcat gccgc当地 ggccaa actggccggcc aaaccggatc cacagc当地 ctg atgttaag 13860

acggcaccaa cgactggaac gccctacacc aggatctcta cggcgagttg gtgttccgc 13920
tgcagggtgt gatcaacctg agcgatccgg aaaccgacta caccggcggc gagttcctgc 13980
ttgtcgaaca gcggcctcgc gcccaatccc ggggtaccgc aatgcaactt ccgcaggac 14040
atggttatgt gttcacgacc cgtgatccggc cggtgcggac tagccgtggc tggtcggcat 14100
ctccagtgcg ccatgggctt tcgactattc gttccggcga acgctatgcc atggggctga 14160
tctttcacga cgacgctga ttgcacgcca tctatagata gcctgtctga ttcaccaatc 14220
gcaccgacga tgccccatcg gcgtagaact cggcgatgct cagcgatgcc agatcaagat 14280
gcaaccgata taggacgccc gacccggcat ccaacgccag ccgcaacaac atttgatcg 14340
gcgtgacatg tgacaccacc agcaccgtcg cgccttcgta gccaacgatg atccgatcac 14400
gtccccgccc aacccggcgc agcacgtcgt cgaagcttc cccacccggg ggcgtgatgc 14460
tggtgtcctg cagccagcga cgggtcagct cgggatcgcg ttctgcggcc tccgcgaacg 14520
tcagccccctc ccaggcgccg aagtccgtct cgaccaggcgt gtcatcgacg accacgtcca 14580
gggccaggc tctggcggcg gtcaccgcgg tgtcgtaagc ccgtctgtac ggcgaggaga 14640
ccaccgcagc gatccccggc cgccgcgcca gataccggc cgccgcacca acctggcgcc 14700
accccacctc gttcaacccc gggttccgc gccccgaata gccgcgttgc tccgacagct 14760
ccgtctgccc gtggcgcaac aaaagttagtc ggggtgggtgt accgcgggccc cgggtccagc 14820
cgggagatgt cgggtactcg gtcgcaacga ttttggcagg atccgcatcc gccgcagccg 14880
attgcgcggc gggtccatc ggttcattgg ccaaccggc tgcatacgtg ttccgggcac 14940
gcggaaacca ctcgtagttg atcctgcgaa actgggacgc caacgcctga gcctggacat 15000
agagcttcag cagatccggg tgcttgaccc tccaccgcgg ggcacatctgc tccaccacca 15060
gcttggagtc catcagcacc gcccgcctgg tggcacctag tttcacggcg tgcgtccaaac 15120
cggttatcag gcccggatat tcggcgacgt tggtcgctgc cggccgatc gcctgttgg 15180
actcggccag cacgggtggag tgatccggcgg tccacaccac cgcgcgtat cggccggcgc 15240
cgggattgcc cccgcgtccg cggcggctt cgtacacaac tttcactcct caaatccttc 15300
gagccgcaac aagatcgctc cgcattccgg gcagcgacc accttcatcct cggccggcgc 15360
cgagatctgg gccagctcg cgcggccgtat ctcgtccgg caggcaccac atcgatgacc 15420
ttgcaaccgc cccgcctcg gcccgcctcc ggccgcgtt ctgtcgatgaaatgttgc gcccggcgttgc 15480
ctcgggatca agtgcgtccg tcagcatgtc gcgtgcgtat gaatgttgc gcccggcttgc 15540
gtcgatttcg gcaagtgcct cgtccaaagc ctgtggcg gcccgcgtat cggccggcgc 15600
cgcttggagc gcccgcgtat cggccgtctg ttgagcctgc agtcctcgc ggcgttccag 15660
cacctccagc agggcatctt ccaaactggc ttgacggcgt tgcaagctgt cgagctcg 15720
ctgcagatca gccaatttgc tggcgccgt tgcacccgaa gtgagcaacg accggccccg 15780

gtcgccacgc ttacgcacccg catcgatctc cgactcaaaa cgcgacaccc ggccgtccaa 15840
gtcctccgccc gcgattcgca gggccgcatt cctgtcggtg gcgccgttgt gctcggcctg 15900
cacctgctgg taagccgccc gctgcggcag atggtagcc cgatgcgcga tccgggtcag 15960
ctcagcatcc agcttcgcca attccagtag cgaccgttgc tgtgccactc cggctttcat 16020
gcctgatctc tcccagtttc gtgatcgagg ttccacgggt cggtgcagat ggtgcacaca 16080
cgcacccggca gcgacgcgcc gaaatgagac cgcaacactt cggcggcctg gccgcaccac 16140
gggaattcgc ttgcccatacg cgcgacgtcg atcaggggca cttgcgaagc tcggcaatgc 16200
tcgtcggctg gatgatgtcg cagatcgccc gtaacgtacg cttgcacgtc cgccggcggcc 16260
acggtgccaa gcaacgagtc cccggcgccg cccgacgaccg cgacccgcga caccacgagg 16320
tcgggatccc cggcggcgcg cacaccggcgc gcaacgcggc ctccagacgg 16380
gcaacaaagg tgccgacgccc ttccgggttt ggcagtctgc caatccggcc taaccggctg 16440
ccgaccggcg gtggtaccag cgcgaaatgc tcgaatgcgg gtcctcgta agggtgccgc 16500
gcccgcacatcg ccccaacac ctcggcgcgc gctcgtgcgg gtgcgacgac ctcgacccgg 16560
tcctcgccca cccgttcgac ggtaccgacg ctgcctatgg cgggcacgc cccgtcgac 16620
gccaggaact gcccggtacc cgcgacactc cagctgcagt gcgagtagtc gccgatatgg 16680
ccggcaccgg cctcaaagac cgctgcccgc accgcctctg agttctcgcg cggcacatag 16740
atgaccacact tgctcagatc ggccgctccg ggcacccgggt cgagaacgcgc gtcgacggc 16800
agaccaacag cgtgtgccag cgctcgac acacccggcg acgcccggc ggcgttggtg 16860
tgccgcgtaa acaacgagcg accggtccgg atcaggccgt gcaccagcac accctttggc 16920
gtgttggccg cgaccgtatc gacccacgc agtaacaacg ggtggtgac caatagcagt 16980
ccggcctggg gaacctggc caccacccgc ggcgtcgct ccaccgcaac ggtcaccgaa 17040
tccaccacgt cgtcggggtc gcccacacc agacccaccc aatcccacga ctggcaagc 17100
cgccggccgggt aggccctggtc cagcacgtcg atgacatcg ccagccgcac actcatcgcc 17160
gtcctccacg ctttgcac ctcggcgtatcg ccgcacccag cacgggcccac tccggcgcga 17220
ccgcccggcc caggtaccgc gcgtccaggc cgacgaagggt gtcaccgcgg cgcacccgcaa 17280
ttccttgct ctgcacatcg tttcgtaatc cgtcagcatc ggcgtatgtt aacagtacga 17340
aaggggccgc accatcgacc acctcgccac ccaccgatct cagtcggcc accatctccg 17400
cgccgcacgc cgtcaacccgc accgcacccgc ctgcggcagc ggcgacccgc cggggggcgc 17460
agcaagcagc gatggccgtc agttgcaatg ttcccaacgg ccagtgcgt cgctgcacgg 17520
tcaaccgagc cagcacgtct ggcgagccga ggcgttagcc cacccgcaat ccggccagcg 17580
accacgtttt cgtcaagcta cggagcacca gcacatcggg cagcgagtc tcggccaacg 17640
attgcggctc gccgggaacc caatcagcga acgcctcgac gaccaccagg atgcgtcccg 17700

gccggcgtaa ctcgagcagc tgctcgccga ggtgcagcac cgaggtgggg ttggtcggat 17760
tacccacgac gacaagggtcg gcgtcgtag gcacgtgcgc ggtgtccagc acgaacggcg 17820
gcttttaggac aacatggtgc gccgtgattc cggcagcgct`caaggctatg gccggctcgg 17880
tgaacgcggg cacgacgatt gctgcccga ccggacttag gttgtgcagc aatgcgaatc 17940
cctccggcgc cccgacgagc gggagcactt cgtcacgggt tctgccatga cgttcagcga 18000
ccgcgtcttgcg acatcgtagg tgctcggata gcgggcccagc tccggcagca 18060
gcccggcgag ctgcccggacc aaccattccg ggggccggtc atggcggacg ttgacggcga 18120
agtccagcac gccggggcgcg acatcctgat caccgtggta gcgcgcggcg gcaagcgggc 18180
tagtgtctag actcgccaca gcgtcaaaca gtagtgggcc ggtgtgcggg ccaagaatcc 18240
agagcaccgcg cgacgcgttg tctacgcggc gacaaccgcg acatcacagg cagctaacag 18300
ggcgtcggcg gtgatgatcg tcaggccaag cagctgtgcc tgggcgatga gcacacggc 18360
aatggatgt cgatggtgat ccggaagctc tgccgtgcgc agtgtgtgcg tggtaactg 18420
acagcggcga cgtgcccgcag cggcgcattc gatcgggcac gtaagaagcc gatggctcg 18480
gcccggggag cttccggagg cggtagttga tcgcgatctc ccaggcactg gcggccgaca 18540
agagaatgct gttgcggacg tcctgaacaa tcgcccgtgt ttcgttgacg gcatccgcag 18600
ccaaacgtgg gtgtcgatga ggtagcgctt caccggtaa acgcgttcgag cacgtcgct 18660
gacaacggag cgtccaaatc gtcgggcacg cggtagcacgc catggtaat gcctaaccgc 18720
cgagtctcat gaggatgcag cggcacaaggc tttgttacgg gtcgcggcg gcgggcaatc 18780
tcaacctctg cccggcgttag acgagccgca gcagctcgga caggcgtgtc ttcgcctcg 18840
gaacgcgcac ccgcttcgca ggcggccaga ctttcgcgtc gaccacgtc tcaccaaact 18900
tcgcgatcat cgccgtatac cacagcgcca acgggttagcg gtttgcctaa ccgcgttcgtc 18960
aacgacaatg ggatcgtagc cgacacgcacc gcgagcggga ccaattgcgg gcctccctca 19020
cgccggcccg cacggcgccgc atcgccgtccg ggtgaatcgcc cgcagctggt gatcttcgt 19080
ctggacggca cgctgaccga ctcggcgccgc ggaatcgat ccagcttccg acacgcgttc 19140
aaccacatcg gtgccccagt acccgaaggc gacgtggcca ctcacatcg tggccggccc 19200
atgcatgaga cgctgcccgc catggggctc ggcgaatccg ccgaggaggc gatcgtagcc 19260
taccggcccg actacagcgcc cccgggttgg gcgatgaaca gtttgcgtca cgggatcg 19320
ccgctgtgg ccgacctgcg caccggccgt gtccggctgg ccgtcgccac ctccaaaggca 19380
gagccgaccg cacggcgaaat cctgcgcac ttcggaaattg agcagcactt cgaggtcatc 19440
gcggggcgca gcaccgatgg ctcgcgaggc agcaagggtcg acgtgtggc ccacgcgttc 19500
gcccggcgccgca ggcggctacc cgagcgggttgcgtatggc ggcggccag ccacgcgtc 19560
gacggggccgg ccgcccgcacgg catcgacacg gttttgcgtca gctggggctaa cgggcccggcc 19620

gacttatcg acaagacctc caccaccgtc gtgacgcatttccgacacat tgacgagctg 19680
agggaggcgc taggtgtctg atccgctgca cgtcacattc gtttgcacgg gcaacatctg 19740
ccggtcgcca atggccgaga agatgttcgc ccaacagctt cggccaccgtg gcctgggtga 19800
cgccgtgcga gtgaccagtg cgggcacccgg gaactggcat gttaggcagg tgcgcacg 19860
gccccggcc ggggtgttc gagccacgg ctaccctacc gaccaccggg ccgcacaagt 19920
cgccaccgaa cacctggcgg cagacctgtt ggtggcattt gaccgcaacc acgctcgct 19980
gttgcggcag ctcggcgtcg aagccgcccgg ggtacggatg ctgcggcat tcgacccacg 20040
ctcgggaacc catgcgctcg atgtcgagga tccctactat ggcatcact ccgacttcga 20100
ggaggtcttc gcotgtcatcg aatccgcctt gcccggcctg cacgactggg tcgacgaacg 20160
tctcgccgg aacggaccga gttgatgccc cgcctagcgt tcctgctgcg gcccggctgg 20220
ctggcgttgg ccctggcgtt ggtcgccgtt acctacctgt gctttacggt gctcgccgg 20280
tggcagctgg gcaagaatgc caaaacgtca cgagagaacc agcagatcag gtattccctc 20340
gacaccccgcc cggttcccgct gaaaaccctt ctaccacagc aggattcgtc ggccggac 20400
gcccggcggc gcccgggtgac ggcaaccggc cagttacccgtt cggacgtgca ggtgctggcc 20460
cgactgcgcg tgggggggggg ggaccaggcg tttgagggtgt tggcccccatt cgtggcgtac 20520
ggcggaccaa ccgtccttggt cgaccgttga tacgtcgccg cccaggtggg ctgcacgtt 20580
ccaccgatcc cccgcctgccc ggtcgacgatcg gtgaccatca ccgcgcggct gcgtgactcc 20640
gaaccgagcg tggggggcaaa agacccattt gtcagagacg gttccagca ggtgtattcg 20700
atcaataccg gacaggtcgc cgcgttgcacc gggtccacgc tggctgggtc ctatctgcag 20760
ttgatcgaag accaaccggc cgggctcgcc gtgctggcg ttccgcattt agatccggg 20820
ccgttcctgt cctatggcat ccaatggatc tcgttcggca ttctggcacc gatcggttgc 20880
ggctatttcg cctacgcccga gatccggcg cggccgggg aaaaagcggg gtcgcccacca 20940
ccggacaaggc caatgacggt cgagcagaaa ctcgttgcacc gctacggccg cggccggtaa 21000
accaacatca cggccaatac cgcagccccc gcctggacca cccgcacag caccacggcg 21060
cgccgcacat cggccacccctt gggcgaccgg ccgtcgccca aggtggcccg gatctgcac 21120
tcatgggtgtt accgggtggg cccacccacgc cgcacgttca ggcggccacgc aaacggcc 21180
tcgacgacac cggcggtggg gctggatgg cggcgccgt cgcgcgcacca gggccgtacc 21240
gcaccggccgg gcgacccacc gaccaccggc ggcacatca ccaccacgc cggccgtcc 21300
cgtgcgcacca catagttggc ccagtcatcc aatctgtgtt cagcccaacc gaatggaga 21360
taacgcggcg agcggttagcc gatcatcgatcc tccagggtgt tgatggcacc atatccac 21420
accgcaggca cggccgttca gacccggccac agcagcggca ccacccgttgc gtcggccgt 21480
ttttcggccca cggactccag cggccgttca gtcaggcccg ggccggccacccag ctggggccggg 21540

tcacgcccgc acagcgacgg cagcagccgt cgccgcgcct cgacatcg tc ggcctccaac 21600
aggtccgata tctggcggcc ggtgcgcgcc agcgaagttc cgcccgacgc tgcccagggtg 21660
gccgtcgccgg tggccgcccac gggccaggac ctgcccggta gccgtcgac tgccgcgcgg 21720
agcaaggccca ccgcgcgcac cagcaggccg acgtgtaccg caccggcgcac ccggccgtca 21780
cggttaggtga tctgtccag ctggcggcc gcccggccga acaggggccac cggtatgaccc 21840
cgtttggggc cggccgaacac gacgtcgagc aggccggccga tcagcacgccc gacggccctg 21900
gtctgcccagg tcgatgcaaa cactccggca gcgtcgacaca cgtggtctac gctcaagctat 21960
ttatgaccc tc atacggcagc tatccacgat gaagcggcca gctacccggg ttggccgaccc 22020
gttgaaccccg gcccgaatgt tggaaagccc ggtggacagc ggcaacgtct acaagcatcc 22080
gttcaagcgg gcccggacca ccggcaccta cctggcgggtg ggcaccatcg ggacggaatc 22200
cgaccgagcg ctgatccggg gtggcggtga cgtcgccac ccggcagggttc ggtcgacggc 22260
ctcgagccca gtgtccata acgccttcga cccgaagttt cagctgtggg tggcggcgtg 22320
tctgtaccgc tacttcgtgg accagcacga gtttctgtac ggcccactcg aagatgccac 22380
cgccgacgccc gtctaccaag acgccaacg gtttagggacc acgctgcagg tgccggaggg 22440
gatgtggccg ccggaccggg tcgcgttcga cgagttactgg aagcgctcg tc ttgatgggct 22500
gcagatcgac gcggccggcgc ggcggcatct tcgcgggggtg gcctcggtag cgtttctccc 22560
gtggccgttg cgccgggtgg ccggggccgtt caacctgttt ggcggacacgg gattcttggc 22620
accggagttc cgccgcgtga tgcagcttgg a gttgtcacag gcccggcgc gtcgttcga 22680
gtggttactt tccgtctac gtttagccga ccggctgttccgc catcgccggg cctggatctt 22740
cgtttaccag cttaacttgcg gggacatgcg gtttcgcgcgc cgacacggcc gccgaatcg 22800
ctgatagagc ccggccgagt gtgagcctga cagccgacca ccggccggcgt gtgtcgccgc 22860
gccaggttca cgctcgccga tctagagccg ccgaaaaccc acttctgggt tgcctcccg 22920
atcaacgtgc tgatctgctc gagcagctca cgcatatcg cgccatcgatcg atccaccgc 22980
gcatacaggt cggcccttggc cgccggcgcg tggccgcacg tcattggccg caccggccgt 23040
gctgtctgcg cgcggccgcgt gtgcgtttga aaccctggcgtc gtcacccac gaccacgaca 23100
ctgcccatacc cggccggcccg ccgacaacga agcacagcta gcccggggc gccggacggga 23160
tcgaaccggcc gaccgtgtt gttgtaaaacc agagctctac cgctgagcta cgcccccattg 23220
accggccgcag gctacacgcg ttgcggccaa gcacccaaaa ccttagggccg taagcgccgc 23280
cagagcgctcg gtccacagcc gctgatcgac aacttcaccc ggctgcttca tctcggcga 23340
ccgaatgatc cctgaccgat cgaccacaaa ggtgccccggg tttagcgatgc cggcctgctc 23400
gttgaagacg ccgttaggcct gactgaccgc gccgtgtggc cagaagtccg acaacacgcgg 23460

aaacgtgaat ccgctctgct tcgcccagat ctttgagtg ggtggcgggc ccaccogaaat 23520
cgctagcgcg gcgcgtcgct cgttctcaaa ctcgggcagg tgcacgcactggcc 23580
ctcgccctgg cagatgcccgg tgaacccaa cggaaagaac accaacagca cgttcttgc 23640
accccggttag cgcgcaggg tgacaagctg ctgattctgg tcgcgcaacg tgaagtcaagg 23700
ggcggtggct cgcacgttca gcatcagcgc ttgcagcccc gcgatttcgg ctgtaccaat 23760
ctgctggcgc tccagttgcc cagattgacc gacgaggctg gcatcagccc agctgtggc 23820
gccgcctcgg caatctcggc gggcaataca tggccgggtt ggccggtctt gggcgtcacc 23880
acccaaatca caccgtcctc ggcgagcggg ccgatcgcat ccatcagggt gtccaccaaa 23940
tcgcccgtcgc catcagcaca ccacaacagg acgacatcga tgacccgtc ggtgtcttca 24000
tcgagcaact ctcccccga cgcgttcccg atggccgcgc ggatgtcgct gtcgggtct 24060
tcgtcccaagc cccattccctg gataagttgg ttcgttggta tgcccaattt gcggcgttag 24120
ttcgaggcgt gatccgcgc gaccaccgtg gaacctcctt cagtctccgc gggccatgtg 24180
cacaccgtcg cgatggcat tatcgctcgca cagccagaac cggtccaccc gcccgcctca 24240
gaaggcggcc acgcacattt tcaatgcctt tgtcttggta tcgttggcc gatcaacccg 24300
ccgggttaat tccgctgtcg acgcgtcgcc accgatggca ttggccaccc cgccggccgc 24360
gtcgacatat gcgttggcgcat cccatcccgat ttgcgcggac agcgcggcgc tcagactgcc 24420
tgagaccgtc gaggcactgt ttttggcgcat gtcgttggcc ggacccgttggta tcggcccggt 24480
gttgcggccc tgattgaacg cggccacgtt ggcgttccacc ttgtcgatgg cgtccttgc 24540
ggtgtggcc acgcgcgtcac acgaggtcgat aatgccttgc gtcgttggcc gatgttggcg 24600
ctgcgactcc cggatgtcgat acgtcgccgc cgaagccgac accgacgcgg acaccgacga 24660
gcggtagggcc ggtgcgcacgt tgggtcggtt catggccgtt ccgtcggttgc cagtggta 24720
tccgacgatc cccatcagca gcagccgcgtt gcagccgagc gccaggccgc ctcgcctgg 24780
gagctcccccc ccgtgcctgc gaggcacggc ggcgcgttgc atgagcacgg catgttggat 24840
tacctggtcg cagcgcgacc ggcgtggccg tgggtgtcgat cgcacccgcgaaaccgagcg 24900
gagtgcggct atccgcgcgc gacgcgggtt cggcacgata gggggacgac catctaaaca 24960
gcacgcacgc ggaagccgcgac cacctacagg agtagtgcgt tgaccacccgat tttcgccgc 25020
cacgatctgg cccaaaactc aaacagcgca agcgaacccg accgagttcg ggtgtccgc 25080
gagggtgtgg cgtcgatattt gcccgcattt gatcccgagg agacccgttgc gttggctgg 25140
tcctttgaca cgcgtcgatc acgctgcggc cgcgtcgccg cccgcgttgc gatgttgcgg 25200
ctgcttagacgc gggccggcga gcagcgggtt ggcgcgttgc cattgcgtc taccgactat 25260
gtcaacacca tcccgaccga gctggagccg tgggtccccgg ggcgcacgaaat cgtcgacgt 25320
cgttatcgag cgtggatcgat atggaaatcgat ggcacccgttgc gcaacgcaccg 25380

ggtgtggcg tgggtggcca tatctcgacc tacgcgtcg ctgcggcgct ctatgaggtc 25440
ggtttcaacc acttcttccg cggcaagtcg caccggcg gcggcgatca ggtgttcatc 25500
cagggccacg ctccccggg aatctaogcg cgcccttcc tcgaaggcg gttgaccgccc 25560
gagcaactcg acggattccg ccaggaacac agccatgtcg gcggcggtt gccgtcctat 25620
ccgcacccgc ggctcatgcc cgacttctgg gaatccccca ccgtgtcgat gggtttggc 25680
ccgctcaacg ccatctacca ggcacggttc aaccactatc tgcatgaccg cggtatcaaa 25740
gacaccccg atcaaacacgt gtgggtttt ttgggcgacg gcgagatgga cgaacccgag 25800
agccgtggc tggcccacgt cggcgcgtg gaaggcttgg acaacttgc cttcgtgatc 25860
aactgcaatc tgcagcgact cgacggcccg gtgcgcggca acggcaagat catccaggag 25920
ctggagtcgt tcttccgcgg tgccggctgg aacgtcatca aggtggtgc gggccgcgaa 25980
tgggatgccc tgctgcacgc cgaccgcgac ggtgcgcgtt gtaatttaat gaataacaaca 26040
cccgatggcg attaccagac ctataaggcc aacgacggcg gctacgtgcg tgaccacttc 26100
ttcggccgcg acccacgcac caaggcgctg gtggagaaca tgagcgacca ggatatctgg 26160
aacctcaaac gggcgccca cgattaccgc aaggtttacg ccgccttaccg cgccgcgc 26220
gaccacaagg gacagccgac ggtgatccgt gccaagacca tcaaaggcta cgcgctggc 26280
aagcatttcg aaggacgc当地 tgccaccac cagatgaaaa aactgaccct ggaagaccc 26340
aaggagtttgc tgacacgc当地 gcggattccg gtcagcgacg cccagcttgc agagaatccg 26400
tacctgccgc cctactacca cccccgc当地 aacgccccgg agattcgta catgctcgac 26460
cgccgcgggg ccctcggggg ctttgttccc gagcgcagga ccaagtccaa agcgctgacc 26520
ctgcccggc当地 ggcacatcta cgccgcgtg aaaaagggtt ctgggcacca ggaggtggcc 26580
accacccatgg cgacgggtgc当地 cacgttccaa gaagtgttgc ggcacaaagca gatcggccg 26640
cgatagtc当地 cgatcattcc cgacgaggcc cgacccctcg ggatggactc ctggttcccg 26700
tcgctaaaga tctataaccg caatggccag ctgtataccg cgggtgacgc cgacctgatc 26760
ctggcctaca aggagagc当地 agtcgggc当地 atcctgcacg agggcatcaa cgaagccggg 26820
tcgggtggct cggtcatcgcc ggccggcacc tcgtatgc当地 cgacacaacgc accgatgatc 26880
cccatattaca tcttctactc gatgttccg当地 ttccagcgca ccggcgatag ctctggcc 26940
gcggccgacc agatggctcg当地 agggttcg当地 ctcggggcc当地 cgcggggcg caccaccctg 27000
accgggtgagg gcctgcaaca cgccgacggc当地 cactcggtc tgctggccgc caccaaccc 27060
gcgggtggcc当地 cctacgaccc ggccctcgcc tacgaaatcg cctacatcg当地 ggaaagcgga 27120
ctggccagga tgtgcgggg当地 gaacccggag aacatcttct tctacatcac cgtctacaac 27180
gagccgtacg tgcaaggcc当地 ggagccggag aacttcgatc cgcggggcg gtcgtgggg 27240
atctaccgct atcacgc当地 caccgagcaa cgcaccaaca aggccgacat cctggccctcc 27300

ggggtagcga tgcccgccgc gctgcggca gcacagatgc tggccgcga gtggatgtc 27360
gccgccgacg tggatcggt gaccagtgg ggcgagctaa accgcgacgg ggtggccatc 27420
gagaccgaga agctccgcca ccccgatcg cgccgcggcg tgcctacgt gacgagagcg 27480
ctggagaatg ctcggggccc ggtgatcgcg gtgtcgact ggatgcgcgc ggtccccgag 27540
cagatccgac cgtgggtgcc gggcacatac ctcacgttgg gcaccgacgg gttcggctt 27600
tccgacactc ggccccccgc tcgcccgtac ttcaacaccc acgcccgaatc ccaggtggtc 27660
gcggttttgg aggcgttggc gggcgacggc gagatcgacc catcggtgcc ggtcggccgc 27720
gcccgccagt accggatcga cgacgtggcg gctgcgcggc agcagaccac ggatcccggt 27780
cccgccccct aacgcggcg agccgaccgc ctggccga atcttccaga aatctggcgt 27840
agcttttagg agtgaacgac aatcagttgg ctccagttgc cggcccgagg tgcgcgtcg 27900
aactgctgga caotgtgccc gattcgctgc tgcggcggtt gaagcagtac tggggccggc 27960
tggccaccga ggcagtttcg gccatgcaag aacggttgcc gttttcgcc gacctagaag 28020
cggtccagcg cgccagcgtg ggcgtggcg tgcagacggc cgtggtaac ttgcgtcaat 28080
ggatgcacga cccgcacagt gacgtcgct ataccgcga ggcattcgag ctggtgcggc 28140
aggatctgac ggcacggatc ggcgtgcgc agaccgtggc catggtgcgg gtcaccatgg 28200
agttcttcga agaagtcgtg cccctgctcg cccgttccga agagcagttt accgcctca 28260
cggtggcat tttgaaatac agccgcgacc tggcattcac cgccgccaacg gcctacgccc 28320
atgcggccga ggcacgaggc acctgggaca gccgatggc ggccagcgtg gtggacgcgg 28380
tggtacgcgg cgacaccgtt cccgagctgc tgcggccggc ggccgcgtt aattgggaca 28440
ccaccgcgccc ggcgaccgta ctggggaa ctccggccgc cggtaaaat ggctccaaca 28500
gcgcacggcga cagcgacggg gccagccagg atgtccgcga caccgcggct cgccacggcc 28560
gcgcgtcgct gaccgacgtg cacggcacct ggctggcgtc gatcgctcc ggccagctgt 28620
cgccaaccga gaagttcctc aaagacctgc tggcagcatt cgccgacgcgc cccgtggta 28680
tcggcccccac ggcgccccatg ctgaccgcgg cgccacccgag cgctagcgtc gcatctccg 28740
ggatgaacgc cgtcggccgc tggcgcggag cggccggcc cgtgcgtggct agggaaacttt 28800
tgcccgaacg cgccctgatg ggcgacgcct cggcgatcg tgcgttcgtt accgacgtga 28860
tgcggcccccct agccgatgcc ggaccgacgc tcatcgagac gctagacgcata tatctggatt 28920
gtggcggccgc gattgaagct tgcgttcgtt tgcgttcgtt acgtgcgtt 28980
accggctcaa gcgatcacc gacttcaccc ggcgcgatcc caccgcgcgc cgcgtgcct 29040
atgtccctcg ggtggcgcc accgtgggtc aactcaacta tccgacgcgcgc cactgaagca 29100
tcgacagcaa tgccgtgtca tagattccct cggccgtcag agggggtcca gcaggggccc 29160
cgaaaagata ccaggggcgc cgtcgacgg aaagtgtatcc agacaacagg tcgcggacgc 29220

atctcaaaaa catagcttac aggcccgtt tgggttat atacaaaaac ctaagacgag 29280
gttcataatc ttttacaccg cgcaaaaccg tcttcacagt gttctcttag acacgtgatt 29340
gggttgcctg caccggaca ggttcgaa accgaggaa tgggtcgcc gtggcttcag 29400
ctgcccggcg cagggacca gatcgcggcg tggcgaaag ccgctgatct agatcttgcc 29460
cggttggca ccacccctc gaccgaggag atcaccgaca ccgcgtcgc ccagccattg 29520
atcgctcccg cgactctgtt ggcccaccag gaactggcgc gccgatgcgt gctcgccggc 29580
aaggacgtca tctggccgg ccactccgtc ggcgaaatcg ccgcctacgc aatcgccgg 29640
gtgatagccg ccgacgacgc cgtcgcgctg gcccaccc gggcgccga gatggccaag 29700
gcctgcgcca cggagccgac cggcatgtt ggggtcgctg gggcgacga gaccgaggt 29760
ctgagtcgcc tggcgagct cgacttggtc cggcaaaacc gcaacgccc cggccagatc 29820
gtcgctcccg gccggctgac cgcgttggag aagctcgccg aagacccggc gccaaggcg 29880
cggtgcgtg cactgggtgt cggcgagcg ttccacaccg agttcatggc gcccgcactt 29940
gacggctttg cggcgccgc gccaacatc gcaaccgccc accccaccgc cacgctgctg 30000
tccaaccgcg acggaaagcc ggtgacatcc gggccgcgg cgatggacac cctggctcc 30060
cagctcaccc aaccggtgcg atggacactg tgcaccgcga cgctgcgcga acacacagtc 30120
acggcgatcg tggagttccc ccccgccggc acgcttagcg gtatcgccaa acgcaactt 30180
cggggggttc cggcacgcgc cgtcaagtca cccgcagacc tggacgagct ggcaaaccta 30240
taaccgcgga ctggccaga acaaccacat acccgctagt tcgatttta cacaacatat 30300
tacgaaggga agcatgtgt gcctgtact caggaagaaa tcattggcg tatcgccgag 30360
atcatcgaaag aggtAACCGG tatcgagccg tccgagatca ccccgagaa gtcgtcgctc 30420
gacgacctgg acatcgactc gctgtcgatg gtcgagatcg ccgtgcagac cgaggacaag 30480
tacggcgatca agatccccga cgaggacactc gccggctctgc gtaccgtcg tgacgttgc 30540
gcctacatcc agaagctcgaa ggaagaaaac cggaggccgg ctcaggcggt ggcgcgaaag 30600
attgagtcgg agaaccggc tgccgttgc aacgttcagg cgaggcttga ggccgagtcc 30660
aagtgagtcgaa gcttccacc gctaattggcg gtttccccag ctttgggtt accgcccgtca 30720
cagcgacgac gtcgtatctcg cggacatcg agacacgtg gaagggtctg ttggccggcg 30780
agacggccat ccacgcactc gaagacgagt tcgtcaccgaa gtggatcta ggcgtcaaga 30840
tcggcggtca cctcaaggat cggcgatca gacacatggg ccgactcgac atgcgacgca 30900
tgtcgatcg ccacggatg ggcaagttgc tggcgccaca gctatggag tccgcggca 30960
gcccggaggt cgatccagac cggatcgccg ttgttgcgg caccggctta ggtggagccg 31020
agaggattgt cgagagctac gacgtatgaa atgcggccgg ccccccggaaag gtgtccccgc 31080
tggccgttca gatgtatcg cccaaacgggt cggcgccgg gatcggtctg cagttgggg 31140

cccgcgcgg ggtgatgacc ccgggtgtcg cctgttcgtc gggctcgaa gcgatcgccc 31200
 acgcgtggcg tcagatcgtg atgggcgacg ccgacgtcgc cgtctgcggc ggtgtcgaag 31260
 gaccatcgaa ggcgtgccc atcgccggt tctccatgt gcgggccatg tcgacccgca 31320
 acgacgagcc tgagcgggcc tcccgccgt tcgacaagga ccgcgacggc tttgtttcg 31380
 gcgaggccgg tgcgctgatg ctcatcgaga cggaggagca cgccaaagcc cgtggcgcca 31440
 agccgttggc ccgattgctg ggtgccggta tcacctcgga cgcctttcat atggtggcgc 31500
 cccggccga tggtgttgcgt gccggtaggg cgatgactcg ctcgctggag ctggccgggt 31560
 tgcgcggcgc ggacatcgac cacgtcaacg cgacggcac ggcgacgcct atcgccgacg 31620
 ccgcggaggc caacgcattc cgcgccgtc gttgtatca ggccgcgggt tacgcgccc 31680
 agtctgcgtt gggccactcg atcggcgccg tcggtgctcg ctgcggtgc ctcacgggtc 31740
 tgacgctgctg cgacggcgac atccgcgca ccctgaacta cgagacaccc gatcccgaga 31800
 tcgacottga cgtcgccgc ggcgaaccgc gctatggcga ttaccgtac gcagtcaaca 31860
 actcggttgcg gttcggcgcc cacaatgtgg cgcttgcctt cggcggttac tgaagcacga 31920
 catcgccggc cgcgaggccc gaggtggggg tccccccgt tgcggggcg agtcggaccg 31980
 atatgaaagg aacgttgcga agaccaatga cggagctgg taccggaaa gcctttccct 32040
 acgttagtcgt caccggcatc gccatgacga ccgcgtcgcc gaccgacgcg gagactacgt 32100
 ggaagttgtt gctggaccgc caaagcgaaa tccgtacgt cgatgacccca ttgcgtcgagg 32160
 agttcgacct gcaagttcgc atcggcgac atctgttga ggaattcgac caccagctga 32220
 cgcggatcga actgcgcggg atgggataacc tgcagcggtat gtcgcgttgc ctgagccggc 32280
 gcctgtggga aaatgcgcggc tcacccgagg tggacaccaa tcgattgtat gtgtccatcg 32340
 gcaccggcct gggttcgcc gaggactgg tcttcagttt cgcgcgtatgc cgcgctcgcc 32400
 gaatgaaggc ggtctcgccg ctgaccgtgc agaagtacat gccaacggg gcccgcgcgg 32460
 cggtcgggtt ggaacggcac gccaaggccg gggtgatgac gcccgtatcg gcgtgcgcac 32520
 ccggcgccga ggccatcgcc cgtcggtggc agcagattgt gctggagag gccgatgccc 32580
 ccatctgcgg cggcggtggag accaggatcg aagcggtgcc catgcgcggg ttgcgtcaga 32640
 tgcgcgtatcgatgttccacc aacaacgacg accccgcggg tgcgtcgccg ccattcgaca 32700
 gggaccgcga cggcttgcgtt gtcggcgagg gcccgtatcg tctgttgcgtt gagaccgagg 32760
 agcacgcac ggcacgtggc gccaacatcc tggccggat catgggcgcg agcatcacct 32820
 ccgatggctt ccacatgggt gccccggacc ccaacggggaa acgcgcgggg catgcgatta 32880
 cgcggcgat tcagctggcg ggcctcgccc cggcgacat cgaccacgtc aatgcgcacg 32940
 ccaccggcacc ccagggtcgcc gacctggccg aaggcagggc catcaacaac gccttggcg 33000
 gcaaccgacc ggccgtgtac gcccccaagt ctgcgcgttgc ccactcggtg ggccgcggc 33060

gcccggtcga atcgatcttgcg acggatcg cgttgcgcga tcaggtgatc cccggccac 33120
tgaatctggtaaacctcgat cccgagatcg attggacgt ggtggcgggt gaaccgcac 33180
ccggcaatttcccgatccgtatcgatcaataact ctttcggatt cggcggccac aacgtggcaa 33240
tcgccttcgg acggtaactaa accccagcgt tacgcacag gagacctgctgatgacaatca 33300
tggcccccggcggcggttggc gagtcgtcg acccccgca tccgctgttgcggctgagca 33360
acttcttcga cgacggcgcgttggaaattgc tgacgcgacg tgaccgctcc ggagtgcgtgg 33420
ccggccggccggg caccgtcaac ggtgtgcgca ccatcgctt ctgcaccgac ggcaccgtga 33480
tgggcggcgcgcatggcgatc gaggggtgca cgacatcgatcaacgcctac gacactgcca 33540
tcgaagacca gagtcccatttgcgttggccatcttgcatttgcgg tggtgcccgctggctgaag 33600
gtgtgcgggcgctgcacgcgttggccagg tggcgatcgatccgcgatcatccgcgctggctgg 33660
acatcccgca gatctcggttgcgttggcccgccggcggccgttggaccgg 33720
cgttgaccga cgtcgatcgatccgcgttggcccgaaagccgggttggcgatccgcggccacg 33780
tggtgcccgatcgatccgcgttggcccgccggcggccgttggaccgg 33840
accacaagaa gtccgggggtgttggccatcgatccgcgttggcccgccggcggccgttggaccgg 33900
gtggccggccgttggccatcgatccgcgttggcccgccggcggccgttggaccgg 33960
aggccgggtga caccgcacatc caccgcgttgcgttggcccgccggcggccgttggaccgg 34020
tgcgtcccgatcgatccgcgttggcccgccggcggccgttggaccgg 34080
atggccggccgttggccatcgatccgcgttggcccgccggcggccgttggaccgg 34140
tggccaaacaa cccgctacgcgttggcccgccggcggccgttggaccgg 34200
cgcgtttcgatcgatccgcgttggcccgccggcggccgttggaccgg 34260
cgggctatcttgcgttggccatcgatccgcgttggcccgccggcggccgttggaccgg 34320
tgctgcacgcgttggccatcgatccgcgttggcccgccggcggccgttggaccgg 34380
acggccggggcgttggccatcgatccgcgttggcccgccggcggccgttggaccgg 34440
ggccggacgcgttggccatcgatccgcgttggcccgccggcggccgttggaccgg 34500
agaagttggcgttggccatcgatccgcgttggcccgccggcggccgttggaccgg 34560
agcatgagcgatcgatccgcgttggccatcgatccgcgttggcccgccggcggccgttggaccgg 34620
agatcgacccggccatcgatccgcgttggccatcgatccgcgttggcccgccggcggccgttggaccgg 34680
ggccggccgttggccatcgatccgcgttggcccgccggcggccgttggaccgg 34740
acgcgcgaggatcgatccgcgttggccatcgatccgcgttggcccgccggcggccgttggaccgg 34800
tggtcaacgcgttggccatcgatccgcgttggccatcgatccgcgttggccatcgatccgcgttggaccgg 34860
tcgagaaggccatcgatccgcgttggccatcgatccgcgttggccatcgatccgcgttggaccgg 34920
agctcgacgttggccatcgatccgcgttggccatcgatccgcgttggccatcgatccgcgttggaccgg 34980

cacaagagcc ggacctgttggccgcgttgg cggctctcgatgtgttcatggggggcacg 35040
ccgtcgcgctgtctgggcaccgcgcgttgtatggattcgacacggaaaacaccgtcg 35100
ctatccatatgctcgatcccggagtaaggtgcggccacggtcggctctatggtccacc 35160
aacgcgtcggtgccggctcaacgggtgtcaggctcgctcgcgaccgcgcgcgc 35220
ctggcggtggaaatggcgcacga cagttgcgcgcgcgcgcgcgcgcgcgcgc 35280
caactacggtaatgcgtcgcgctcgactgaaattgaaaaa cgcgcgttgcgagcagc 35340
gcgcgcgaggcatcgctcgcgcgcgcgaac tcttacccttcgcgcgcgcgcgc 35400
cgcccatggagacgcgaggctcgctcgatcgaccacggctcgccgttgcgcgc 35460
ttcaatacccgatacacggc caccgtggtaatgtggcg agtcgacttcgcgc 35520
acatgcgtctcgccggccaa tacgaaagca tcgagtgca cgcgggacgcgcggagatgc 35580
tgcgcgacaa gacacgctggcgcacagctcc aagagctcggtggacgattgtcccgatt 35640
tcgtcgacgatgtcagacgcgacccggccgcctggcgccgcgcgcgcgcgc 35700
accgcgcgcgtatggccggctgatcgctggcgcgcgcgcgcgcgcgcgcgc 35760
cgcagtcgatgtcgcgtctcgctcgacccggcgcgcgcgcgcgcgcgcgc 35820
ctacgc 35880
ccacatccagcgccccctca tgggtcaccgcgtattcgaaatccgcgcgcgc 35940
tgccgtcgccgaccggctcggtggccgcctcgatgtggcgcgcgcgcgc 36000
cctcgccccccgtacccgcgcaccagcgcactcgccgcgcgcgcgcgcgc 36060
gc 36120
gcaggtgtcg cagcgtgatgcgcgcgatcagcacatccctctgcgcgcgc 36180
tcagctgccgcgcgaccacaatgcgtcgcgcgcgcgcgcgcgcgcgcgc 36240
gcgaaacgtcgccggccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 36300
atccccataatcccgatgacatccctggtgccgaccgcgcgcgcgcgcgc 36360
ccgtatcccgatcccgatcccgatcccgatcccgatcccgatcccgatcc 36420
cgggtgcgttcgtcgatcccgatcccgatcccgatcccgatcccgatcc 36480
acacgaagcgatcccgatcccgatcccgatcccgatcccgatcccgatcc 36540
acgacttcgcgtcgaagaccatcccgatcccgatcccgatcccgatcccg 36600
cgatctcaccatcccgatcccgatcccgatcccgatcccgatcccgatcc 36660
actgcgcggatcccgatcccgatcccgatcccgatcccgatcccgatcc 36720
cgtaagtggatcccgatcccgatcccgatcccgatcccgatcccgatcc 36780
ggcggtcgatcccgatcccgatcccgatcccgatcccgatcccgatcc 36840
gaacgggtggatcccgatcccgatcccgatcccgatcccgatcccgatcc 36900

tcgcccggcgg cgtacccgct agcacccgca aagcgtcgcc ggccaggaaa ccggcacgca 36960
 ccaacgcccc cttggtgtga cccatcgacg gcaacaacgg gaccagttgc ggcacatggcat 37020
 gcacgagatg aggagcggtt cgtgtcatca ggattcccg cg ttcgacggcg ctgcgcggg 37080
 cgatgcccac gttgccgctg gccagatagc gcagaccgccc gtgcaccaac ttcgagctcc 37140
 agcggctggt gccgaacgccc agatcatgct tttccaccaa ggccaccgtc agaccgcggg 37200
 tggcagcatc taaggcaatg ccaacaccgg taatgccccc gcctatcacg atgacgtcga 37260
 gtgcgccacc gtcggccagt gcggtcaggt cggcggagcg acgcgcgcgg ttgagtgcag 37320
 ccgagtgccc catcagcaca aatatccgtt cagtgcgtgg gtaagttcgg tggccagcgc 37380
 ggccgaatcg aggatcgaat cgacgatgtc cgcggactgg atggtcact gggcgatcg 37440
 caacaccatg gtcggccagt gacgagcgtc gccggagcgc acactgccc accgctgcgc 37500
 cactgtcagc cggccggcca acccctcgat caggacctgc tggctggtgc cgaggcgctc 37560
 ggtgatgtac accctggcca gctccgagtg catgaccgac atgatcagat cgtcaccccg 37620
 caaccggtcg gccaccgcga caatctgctt taccoaacgct tcccggcgtt ccccgatcg 37680
 gggcacttcc cgcagcacgt cggcgatatg gctggtcagc atggacgcca tgatcgaccg 37740
 ggtgtccggc cagcgacggt atacggcgtt gcggctcagc cccgcgcgc gggcgatctc 37800
 ggcaagtgtc acccggtcca cggcgataatc gacgacgcag ctcggcgtt cccgcaggat 37860
 acgaccaccg gtatccgcgc ggtcattact cattgacagc atgtgtataa ctgtaacgcg 37920
 tgactcaccg cgaggaactc cttccacca tgaaatggg cgcgtgggg gatccgcgc 37980
 cggccaagcc actttctgat ggcgtccggt cggtgctgaa gcaggttgc ggcctagcgg 38040
 actcggagca gccccgacgc gaccccgccc aggtgcagct ggcggccgtcc gcccgtcgg 38100
 gggcagacca 38110

<210> 25

<211> 2540

<212> DNA

<213> Homo sapiens

<400> 25
 gaaaagggtgg acaagtccata ttttcaagag aagatgactt ttaacagttt tgaaggatct 60
 aaaacttgt tacctgcaga catcaataag gaagaagaat ttgttagaaga gtttaataga 120
 ttaaaaactt ttgctaattt tccaagtggt agtcctgttt cagcatcaac actggcacga 180
 gcagggtttc ttatatactgg tgaaggagat accgtgcggt gcttttagtt tcatgcagct 240
 gtagatagat ggcaatatgg agactcagca gttggaaagac acaggaaagt atccccaaat 300

| | |
|--|------|
| tgcagattta tcaacggctt ttatcttcaa aatagtgcctt | 360 |
| atccagaatg gtcagtacaa agttgaaaac tatctggaa gcagagatca | 420 |
| gacaggccat ctgagacaca tgcagactat cttttgagaa ctggcaggt tgttagatata | 480 |
| tcagacacca tatacccgag gaaccctgcc atgtattgtt aagaagctt tagaaagtcc | 540 |
| tttcagaact ggcagacta tgctcacca accccaagag agttagcaag tgctggactc | 600 |
| tactacacag gtattggta ccaagtgcag tgctttgtt gtggggaaa actgaaaaat | 660 |
| tgggacacctt gtgatcgtgc ctggcagaa cacaggcgc acctttctaa ttgcttctt | 720 |
| gttttggcc ggaatcttaa tattcgaagt gaatctgtt ctgtgagttc tgataggaat | 780 |
| ttcccaaatt caacaaatct tccaagaaat ccatccatgg cagattatga agcacggatc | 840 |
| tttacttttggat atactcagtt aacaaggagc agcttgcaag agctggattt | 900 |
| tatgctttag gtgaagggtga taaagtaaag tgctttcaact gtggaggagg gctaactgat | 960 |
| tggaaagccca gtgaagaccc ttgggaacaa catgctaaat ggtatccagg gtgcaaataat | 1020 |
| ctgttagaac agaagggaca agaatatata aacaatatttca ttcaacttgc | 1080 |
| gagtgctgg taagaactac tgagaaaaca ccatcactaa ctagaagaat tgatgatacc | 1140 |
| atcttccaaa atcctatggt acaagaagct atacgaatgg gtttcagttt caaggacatt | 1200 |
| aagaaaataa tggagggaaa aattcagata tctggagca actataaatttca acttgaggtt | 1260 |
| ctgggtgcag atcttagtcaa tgctcagaaa gacagtatgc aagatgagtc aagtcaagact | 1320 |
| tcattacaga aagagattt tactgaagag cagctaaggc gcctgcaaga ggagaagctt | 1380 |
| tgcaaaatct gtatggatag aaatattgtt atcgttttttgc ttccctgtgg acatctagtc | 1440 |
| acttgtaaac aatgtgctga agcagttgac aagtgtccca tgtgctacac agtcattact | 1500 |
| ttcaagcaaa aaatttttat gtcttaatct aactctatag taggcattttt atgttgc | 1560 |
| tattacocgtt attgaatgtt tgatgtgaac tgactttaag taatcaggat tgaattccat | 1620 |
| tagcatttgc taccaagtag gaaaaaaaaat gtacatggca gtgttttagt tggcaatata | 1680 |
| atctttgaat ttcttgattt ttcagggtat tagctgttattt atccattttt tttactgttta | 1740 |
| ttaattgaa accatagact aagaataaga agcatcatac tataactgaa cacaatgtgt | 1800 |
| attcatagta tactgatttta atttctaattt gtaagtgaat taatcatctg gattttttat | 1860 |
| tcttttcaaa taggcttaac aaatggagct ttctgtatataa atgtggag attagagttt | 1920 |
| atctcccaa tcacataattt tgttttgtgt gaaaaaggaa taaatttttc catgctgggt | 1980 |
| gaaagataga gattttttt agaggttggat tggtgtttt taggattctg tccattttct | 2040 |
| tgtaaaggaa taaacacggaa cgtgtgcgaa atatgtttgtt aaagtgtt gcccatttttg | 2100 |
| aaagcgtatt taatgtataga atactatcga gccaacatgtt actgacatgg aaagatgtca | 2160 |
| gagatatgtt aagtgtaaaa tgcaagtggc gggacactat gtatgtctg agccagatca | 2220 |

aagtatgtat gttgttaata tgcataagaac gagagatgg gaaagatata caccaaactg 2280
ttaaatgtgg tttcttgcg gggagggggg gattggggga ggggccccag aggggtttta 2340
gaggggcctt ttcaacttgcg actttttca ttttgttctg ttccgatttt ttataagtat 2400
gtagaccccg aagggtttta tgggaactaa catcagtaac ctaaccccg tgactatcct 2460
gtgctttcc tagggagctg tgttgttcc caccaccac cttccctct gaacaaatgc 2520
ctgagtgctg gggcactttg 2540

<210> 26

<211> 103

<212> RNA

<213> Homo sapiens

<400> 26

agcuccuaua acaaaagucu guugcuugug uuucacauuu uggauuuuccu aauauaaugu 60
ucucuuuuua gaaaaggugg acaaguccua uuuucaagag aag 103

<210> 27

<211> 28

<212> RNA

<213> Homo sapiens

<400> 27

ggauuuuccua auauaauguu cucuuuuu 28

<210> 28

<211> 1619

<212> DNA

<213> Homo sapiens

<400> 28

ccgcaggatt tgaatcgccg gaccgttgg cagaggtggc ggccggccca tgggtgcccc 60
gacgttgcgg cctgcctggc agcccttct caaggaccac cgcatctcta cattcaagaa 120
ctggcccttc ttggagggtc gcgcctgcac cccggagcgg atggccgagg ctggcttcatt 180
ccactgcccc actgagaacg agccagactt ggcccaagtgt ttcttctgct tcaaggagct 240
ggaaggctgg gagccagatg acgaccccat agaggaacat aaaaagcatt cgtccgggtt 300

cgctttcctt tctgtcaaga agcagttga agaattaacc cttggtaat ttttgaact 360
ggcagagaa agagccaaga acaaaattgc aaaggaaacc aacaataaga agaaagaatt 420
tgaggaaact gcgaagaaag tgccgtgc catcgacgag ctggctgcca tggattgagg 480
cctctggccg gagotgcctg gtcccagagt ggctgcacca cttccagggt ttattccctg 540
gtgccaccag ctttctgtg ggcccttag caatgtctta ggaaaggaga tcaacat 600
caaatttagat gtttcaactg tgctcctgtt ttgtcttgc aatggcacca gaggtgc 660
tgccctgtca gcgggtgctg ctggtaacag tggctgc tctctctc tctcttttt 720
gggggctcat ttttgcgtt ttgattcccg ggcttaccag gtgagaagtg agggaggaag 780
aaggcagtgt cccttttgc agagctgaca gctttgtcg cgtggcaga gccttcaca 840
gtgaatgtgt ctggacacta tttttgttag gctgtcacag tcctgagtgt ggacttggca 900
ggtcctgtt gaatctgagc tgcaggttcc ttatctgtca cacctgtgcc tcctcagagg 960
acagttttt tttttttttt ggttagatgca tgacttgtt 1020
gtgatgagag aatggagaca gaggccctgg ctccctact gtttacaac atggctttct 1080
tattttgtt gaattgttaa ttcacagaat agcacaact acaattaaaa ctaagcacaa 1140
agccattcta agtcattggg gaaacgggt gaacttcagg tggatgagga gacagaatag 1200
agtgatagga agcgtctggc agatactcct ttgcactg ctgtgtgatt agacaggccc 1260
agtgagccgc gggcacatg ctggccgctc ctccctcaga aaaaggcagt ggcctaaatc 1320
ctttttaat gacttggctc gatgtgtgg gggactggct gggctgctgc aggccgtgt 1380
tctgtcagcc caaccctcac atctgtcacg ttctccacac gggggagaga cgcagtcgc 1440
ccaggtcccc gcttctttg gaggcagcag ctcccgagg gctgaagtct ggcgtaaagat 1500
gatggatttgc attcgccctc ctccctgtca tagagctgca gggtgatttgc ttacagcttc 1560
gctggaaacc tctggaggc atctcggtc ttccctgtca ataaaaagcc tgcatttc 1619

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US02/11758

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) :C12M 1/38, 1/40; C12Q 1/68
US CL :435/6, 91.2, 172.8, 286.1, 286.5, 282.9

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6, 91.2, 172.8, 286.1, 286.5, 282.9

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

WEST: USPAT, DERWENT/EP ABSTRACT.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|-----------|--|-----------------------|
| Y | US 6,060,240 A (KAMB et al.) 09 May 2000, see entire document. | 1 |
| Y | 5,716,825A (HANCOCK et al.) 10 February 1998, see entire document, especially columns 7-8. | 1 |
| A | US 5,667,975 A (DYKSTRA et al.) 16 September 1997, see entire document. | 1 |

 Further documents are listed in the continuation of Box C. See parent family annex.

| | | |
|--|--|-----|
| • Special categories of cited documents: | | |
| "A" | document defining the general state of the art which is not considered to be of particular relevance | "T" |
| "E" | earlier document published on or after the international filing date | "X" |
| "L" | document which may throw doubt on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) | "Y" |
| "O" | document referring to an oral disclosure, use, exhibition or other means | |
| "P" | document published prior to the international filing date but later than the priority date claimed | "E" |

later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

document member of the same patent family

| | |
|---|---|
| Date of the actual completion of the international search | Date of mailing of the international search report |
| 17 JUNE 2002 | 18 SEP 2002 |
| Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 | Authorized officer <i>Valerie Bell-Harris for</i> BENNETT CELSA |
| Facsimile No. (703) 305-3280 | Telephone No. (703) 308-0196 |